Simulating Cryptocurrency Prices using Different Volatility Models

Abstract

In this report, different approaches were considered to simulate price trajectories of cryptocurrencies. Models considered include traditional correlated geometric Brownian motion model and different specifications of multivariate GARCH model: copula-MGARCH, DCC-MGARCH, Student-t MGARCH. All approaches were numerically implemented using real data on the prices of cryptocurrencies supported by the protocol. Traditional Brownian motion model was found to be unable to capture real-data patterns whereas simulation results illustrate rather good performance for MGARCH models. Based on the comparative analysis of computing resources and quality, the copula-MGARCH model was chosen as a candidate for production implementation.

Introduction

Simulation of price trajectories is an important task in the field of both traditional and decentralized finance (DeFi). Simulated paths are broadly used for the purposes of quantitative risk assessment, scenario analysis and sensitivity analysis, finding optimal system parameters, and making management decisions.

To obtain an accurate risk assessment, it is necessary to simulate realistic time series of cryptocurrencies prices that consider the patterns that are present in real observations. Clearly, the task of selecting a proper distribution model for simulating financial returns is still unresolved, there are some generally used approaches and business practices based on the use of both traditional probabilistic models and artificial intelligence models.

According to numerous empirical research the behavior of financial assets highlights some common properties that must be taken into account when developing adequate models of price (return) dynamics. These properties (also called "stylized facts") include volatility clustering – large price changes tend to be followed by large price changes and small price changes tend to be followed by small price changes, fat-tailed distribution of assets' returns – the probability of extreme profits or losses is much larger than predicted by the normal distribution, non-linear dependencies between assets and time-varying correlation structure – correlation may become higher during times of stress. A methodology for prices simulations can be based on any statistically acceptable time-series model capable of capturing these stylized facts.

Different econometric models have been suggested to cover the above features and the most widely used ones are the GARCH-type models. They can describe the time structure of volatility including the volatility clustering. Volatility, or the scale parameter of the distribution, is an important factor in risk estimation because it affects all quantiles of the return distribution including the very extreme ones. Inclusion of a GARCH-type component allows to build realistic volatility models and, as a result, obtain adequate price trajectories.

The objective of this work is to examine different multivariate models for simulation prices of cryptocurrencies that are supported in the Mars protocol and to choose the proper methodology for modelling their joint dynamics. The simulated trajectories will be used in agent-based modeling, the purpose of which is to assess risks and optimize key protocol parameters. We reviewed several specifications of the multivariate GARCH model: copula-MGARCH, DCC-MGARCH, GARCH models with multivariate Student-t innovations (referred to later in this report as Student-t MGARCH model). For comparison we have also presented the results of simulations using the traditional correlated geometric

Brownian motion model, which is still quite often used for simulations of financial data mostly due to its ease of implementation and calculation speed.

Our results illustrate that MGARCH models outperform standard correlated geometric Brownian motion model in terms of their ability to capture real-data features. All considered MGARCH models were found to be of acceptable quality and can be taken as a basis for developing a simulation framework. MGARCH model with multivariate Student-t innovations is the easiest to implement but it is time consuming to generate many paths. DCC-MGARCH model has the advantage of allowing dynamic correlations between cryptocurrencies, but requires additional calibration of the correlation model, which can be problematic when the number of assets in the protocol is large. Tradeoff between the simulation speed and quality is achieved by using the copula-MGARCH model which captures nonlinear dependencies, fat-tails and volatility clusters quite well and allows to quickly generate large samples.

Roadmap. We provide brief technical description of the approaches in Section 1. In Section 2, numerical modelling results are provided and the quality of simulation results is discussed. Finally, we summarize comparative analysis results and future directions of research.

1 Methodology Description

Correlated Brownian Motion

In traditional finance the most widely used model of stock price behavior is the geometric Brownian motion (GBM) model. GBM is given by the following stochastic differential equation (SDE)

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

where S_t is the asset price, μ is the expected return (also called drift) of the asset, σ is the volatility (also called diffusion), W_t is a Wiener process or Brownian motion. The first term in the right-hand side of the SDE is used to model deterministic trends, while the second term is used to model a set of unpredictable events occurring during this motion. The solution of this SDE has the following form:

$$S_{t+\Delta t} = S_t exp\left(\left(\mu - \frac{\sigma^2}{2}\right)\Delta\, t + \sigma\sqrt{\Delta\, t}Z\right), Z \sim \mathcal{N}(0,1)$$

This equation is used to simulate the univariate asset price behavior. However, GBM can be easily extended to the case where there are multiple correlated price paths. Then each price path follows the underlying process

$$dS_t^i = \mu_i S_t^i dt + \sigma_i S_t^i dW_t^i, i = \overline{1, N}$$

where the Wiener processes dW_t^i are correlated such that $E\left(dW_t^i dW_t^j\right) = \rho_{ij}dt$, ρ_{ij} is the correlation coefficient, $\rho_{ii} = 1$. The solution of each SDE has the following form:

$$S_{t+\Delta t}^i = S_t^i exp\left(\left(\mu_i - \frac{\sigma_i^2}{2}\right)\Delta\, t + \sigma_i\sqrt{\Delta\, t}Z_i\right), i = \overline{1, N}.$$

When N correlated assets are modelled simultaneously, at each time step in the simulation N correlated random numbers $Z = [Z_1, \dots, Z_N]^T$ are required. The correlated random numbers are typically calculated using the Cholesky factorization. The following is the procedure for simulation of the correlated price paths based on the GBM that should be repeated at each time step:

1. Estimate expected return μ_i and volatility σ_i of each asset log returns.
2. Estimate correlation matrix of log returns C=[ρ_{ij}] where each ρ_{ij} represents the correlation of the *i*th and *j*th asset. C should be symmetric and positive definite.

3. Perform Cholesky decomposition of the asset correlation matrix $C = LL^T$ into its lower (L) and upper (L^T) triangular parts.
4. Generate a vector of independent Brownian motions $X \sim \mathcal{N}(0, I)$, where I is the $N \times N$ identity matrix.
5. Obtain a vector of correlated Brownian motions $Z = LX$.
6. Simulate correlated prices.

The traditional GBM model is based on the key assumptions:

- the logarithmic change of the stock price is normally distributed;
- volatility (i.e standard deviation of the log returns) is constant;
- expected return (i.e mean of the log returns) is independent of the stock performance.

The above assumptions rarely hold for real data. Particularly, it is known that asset returns have fat tails in the sense that they exhibit more extreme outcomes than a normally distributed random variable with the same mean and variance. Figure below plots the histogram of the 15-minute returns for Ethereum and superimposes the normal distribution with the same mean and variance.

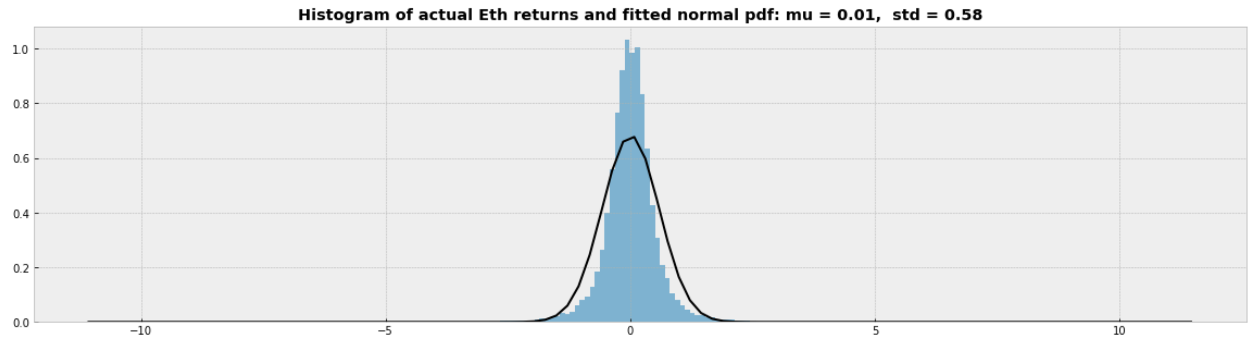

Figure 1. Actual 15-minute log-returns of Ethereum and fitted normal pdf

The following observations can be made:

- the peak of the distribution is much higher than for the normal distribution;
- the sides of the distribution are lower than for the normal distribution;
- the tails of the return distribution are thicker (fatter) than for the normal distribution.

In other words, there are more days when very little happens in the market than predicted by the normal and more days when market prices change considerably.

The left-hand side of Figure 1 displays 15-minute log-returns of prices for Ethereum and Aave over the period from 17/03/21 to 08/12/21. The right-hand side shows correspondingly many simulated data points from a fitted bivariate normal distribution.

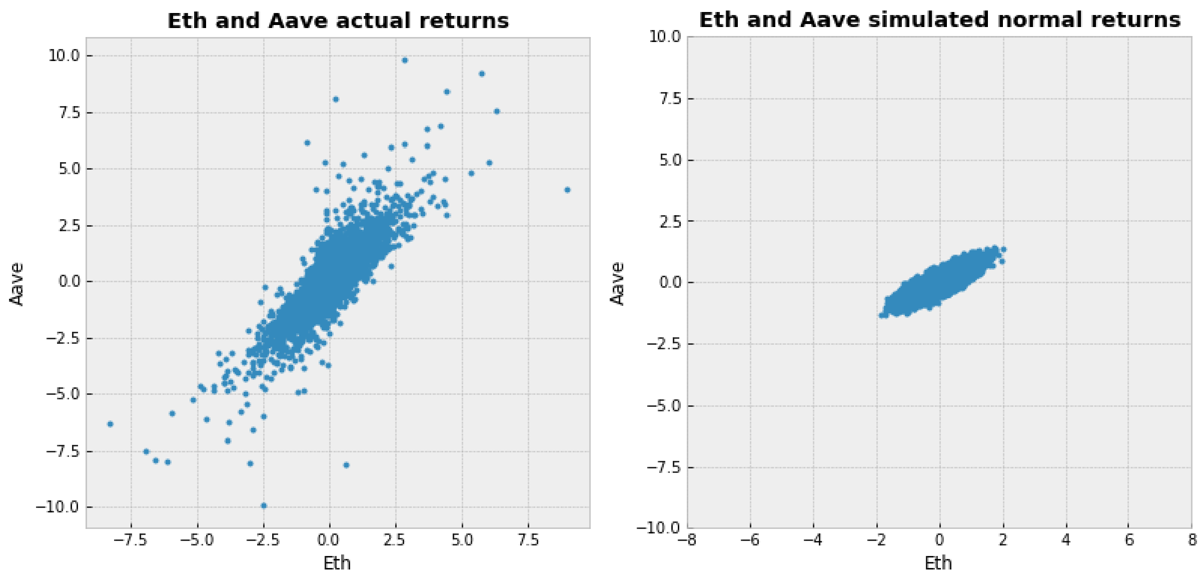

Figure 2. Log-returns for 15-minute (Eth, Aave) data (left) from 17/03/21 to 08/12/21. Correspondingly many simulated data points from a fitted bivariate normal distribution (right)

Figure 2 shows that the bivariate normal distribution is not an adequate model to account for joint "extreme" events (in the upper right or lower left corner of the bivariate distribution).

The assumption on constant volatility is also usually not valid in practice. One of the well-known facts about the returns of financial assets is so-called volatility clustering. It relates to the observation that the magnitudes of the volatilities of returns tend to cluster together, so that we observe many days of high volatility, followed by many days of low volatility. In other words, large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes. The figure below displays 15-minute returns of prices for Ethereum.

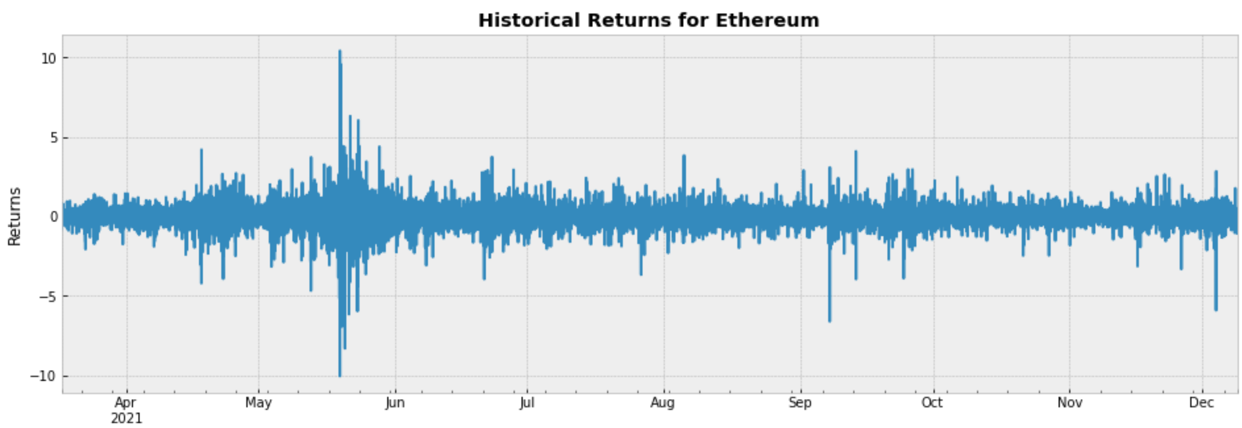

Figure 3. Actual 15-minute returns of Ethereum

It can be seen that actual time-series actually exhibit volatility clusters as there are periods with high and low volatility.

Formal statistical or graphical tests can be made to prove the existence of volatility clusters in data. A standard graphical method for exploring volatility clusters is the autocorrelation function (ACF) of squared returns. The reason for focusing on squared returns is that they are good proxies for volatilities. There are also some well-known statistical tests that can be made, e.g. the Ljung–Box test or the Engle LM test, etc. but we did not conduct them in this study.

MGARCH Models

The majority of time-series models in regular use that allow to capture such real-data phenomena as time-varying volatility and volatility clusters belong to the family of multivariate GARCH models (MGARCH). Currently, a large number of MGARCH-type models for modelling and forecasting volatility exist. The recent survey can be found e.g. in [1] and [3].

For the cryptocurrencies' prices simulation problem, we've chosen several model-candidates from the variety of existing specifications based on their underlying assumptions and complexity of implementation. Particularly, the following MGARCH models were considered:

- Copula-MGARCH model;
- DCC-MGARCH model;
- Univariate GARCH models with multivariate Student-t innovations (Student-t MGARCH).

All considered models initially start with univariate GARCH models for each asset and the methodology for correlation structure is determined separately. Calibration of such models can also be made in two steps, which greatly simplifies the identification of models. When calibrating one-dimensional models, the student-t distribution of innovations is assumed to capture fat-tails of returns.

Clearly more complex GARCH models could be investigated. The goal however is to keep this simple.

The brief description of the above approaches is given in the following sub-sections.

MGARCH Model Methodology

A GARCH model consists of two equations: conditional mean equation, which specifies the behaviour of the returns, and a conditional variance equation that describes the dynamic of conditional volatility. It is generally more efficient to separate estimation of the mean from the volatility estimation.

Note that to include the possibility that returns are autocorrelated, the conditional mean equation could be an autoregressive (AR) model or autoregressive moving average (ARMA) model.

Consider the univariate ARMA(m_i, l_i)-GARCH(p_i, q_i) processes:

$$r_{it} = \mu_{it} + \epsilon_{it}$$

$$\epsilon_{it} = \sigma_{it} Z_{it}$$

where the first subscript indicates the asset and the second subscript the time; $\epsilon_{it} = r_{it} - \mu_{it}$ are called model residuals; $Z_{it} = \frac{\epsilon_{it}}{\sigma_{it}}$ are random variables with zero mean and unit variance, also called standardized residuals or innovations; μ_{it} and σ_{it} are the conditional mean and volatility respectively with the dynamics described by the equations:

$$\mu_{it} = \mu_i + \sum_{k=1}^{m_{i1}} \phi_{ik}(r_{it-k} - \mu_i) + \sum_{k=1}^{l_{i1}} \theta_{ik}(r_{it-k} - \mu_{it-k})$$

$$\sigma_{it}^2 = \alpha_{i0} + \sum_{k=1}^{p_{i2}} \alpha_{ik}(r_{it-k} - \mu_{it-k})^2 + \sum_{k=1}^{q_{i2}} \beta_{ik}\, \sigma_{it-k}^2$$

Note that the mean part can be also modelled via multivariate models, e.g. VARMA models for the conditional mean of several time series taking into account serial dependence between returns, but in that case the number of parameters increases rapidly.

The most common version of GARCH model only employs one lag. Hansen and Lunde [6] published a study where they compared different GARCH models to see if anything outperforms the GARCH (1,1) model. They found no evidence to reject the standard GARCH (1,1) model in favor of other GARCH models. Therefore, to keep the notational burden low, we further present the models in their '(1,1)' form rather than in their general '(p,q)' form and skip the MA-part. The AR(1)-GARCH(1,1) model can be written as:

$$r_{it} = \mu_{it} + \epsilon_{it}$$

$$\mu_{it} = \phi_i + \theta_i r_{it-1}$$

$$\epsilon_{it} = \sigma_{it} z_{it}$$

$$\sigma_{it}^2 = \omega_i + \alpha_i (r_{it-1} - \mu_{it-1})^2 + \beta_i \sigma_{it-1}^2$$

$$\alpha_i + \beta_i < 1$$

Omega (ω_i) is the baseline variance for the model, so the square root of omega is the standard deviation of returns. Alpha (α_i) measures the extent to which a volatility shock today feeds through into next period's volatility. Beta (β_i) is the persistence parameter, if beta is greater than 1 it leads to a positive feedback loop for small shocks that can create runaway volatility. The sum of alpha and beta measures the rate at which the volatility decays and if alpha plus beta equals to 1 then the model has persistent volatility and one might want to look at other models like IGARCH (Integrated GARCH).

We have already seen that the distribution of returns is far from the normal. The GARCH model with assumed conditionally normal distribution might not sufficiently cover the leptokurtosis and fat tails in financial time series. That is why other than Gaussian might be more fitting for innovations in the GARCH model. Some possible choices are: Student-t distribution, Skew-t distribution, Generalized Error Distribution (GED), Skewed Generalized Error Distribution (SGED). In this analysis, we assume a student-t distribution of innovations instead of the normal one.

We fitted a univariate AR(1)-GARCH(1,1) model for each cryptocurrency assuming the Student's t distribution on the residual. The degrees of freedom (DOF) parameter estimated from real data, which governs the tail behavior, is shown in the table below. The smaller the value is, the heavier the tails are. If DOF has a value of about 30 and above, it can be argued that the Student's t distribution is close to the normal distribution.

Cryptocurrency	Student-t DOF
Aave	6.1
Comp	5.8
Eth	5.0

Table 1. Estimated degrees of freedom parameter of the Student-t distribution

We see that all assets have a DOF below 7 indicating a diverse tail behavior.

Now let's consider multidimensional extensions of the GARCH model. Let us introduce the vector notations:

$$r_t = [r_{1t}, \dots, r_{Nt}]^T$$

$$\epsilon_t = [\epsilon_{1t}, \dots, \epsilon_{Nt}]^T$$

$$Z_t = [z_{1t}, \dots, z_{Nt}]^T$$

In matrix form for the most general case of the MGARCH model the innovations can be presented as:

$$\epsilon_t = H_t^{1/2} Z_t$$

where $H_t^{1/2}$ is a $N \times N$ positive definite matrix such that H_t is the conditional variance matrix of r_t, Z_t is the $N \times 1$ random vector with the following first two moments:

$$E\{Z_t\} = 0_{N\times 1}$$

$$V\{Z_t\} = I_N$$

where I_N is the identity matrix of order N.

When returns are modelled via independent univariate GARCH models ignoring correlation, the matrix H_t is simply a diagonal matrix with conditional variances as diagonal elements.

To capture the correlation structure, we need to model the entire covariance matrix including both variances and covariances. The number of volatility terms in the covariance matrix increases more rapidly than the number of assets leading to the problem known as ''the curse of dimensionality''. It complicates the estimation of multivariate volatility models. For that reason, various multivariate volatility models are proposed that allow to parametrize the covariance matrix H_t in order to reduce the number of parameters to be estimated. Examples include VEC, BEKK, OGARCH, CCC-GARCH, DCC-GARCH, copula-GARCH etc. The detailed review can be found in [1] and [3]. Some specifications of the MGARCH models are considered in the following sub-sections.

When the model is estimated by a maximum likelihood method, we can ''break'' the log-likelihood function into two parts: one for parameters determining univariate volatilities and another for parameters determining the correlations. This is known as the two-step estimation technique. Large covariance matrices can be consistently estimated using this technique without requiring too much computational power.

Another problem that often arises in multivariate volatility models is the lack of positive semi-definiteness of a covariance matrix. Unfortunately, ensuring positive semi-definiteness in practical application can be challenging for many, otherwise good, models.

Copula-GARCH

This approach makes use of the theorem due to Sklar (1959) stating that any N-dimensional joint distribution function may be decomposed into its N univariate marginal distributions, and a copula function that completely describes the dependence structure between the N variables. These models are specified by GARCH equations for the conditional variances (possibly with each variance depending on the lag of the other variances and of the other shocks), marginal distributions for each series (e.g. t-distributions) and a conditional copula function.

The dependency structure between the innovations of the GARCH models is modeled by a copula function C, characterized by the vector of parameters.

Let GARCH-innovations follow Student-t marginal distributions with the DOF parameter ν_i:

$$Z_i \sim t(\nu_i)$$

Then by the Sklar's theorem, there is a unique copula C such that the joint cumulative distribution function is defined as:

$$F(z_1, z_2, \dots, z_N) = C(F_1(z_1), F_2(z_2), \dots, F_N(z_N))$$

The copula C contains all information on the dependence structure between the components of $Z_1, \dots, Z_N$. Consider the integral probability transformations:

$$u_i = F_i(z_i)$$

then

$$C(u_1, u_2, \dots, u_N) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_N^{-1}(u_N)\right)$$

The reverse of these steps can be used to generate pseudo-random samples from general classes of multivariate probability distributions. That is, given a procedure to generate a sample $(u_1, u_2, \dots, u_N)$ from the copula function, the required sample can be constructed as

$$(z_1, z_2, \dots, z_N) = (F_1^{-1}(u_1), F_2^{-1}(u_2), \dots, F_N^{-1}(u_N))$$

The most commonly used copulas in finance include: the Gaussian copula, the Student-t copula and the family of Archimedean copulas: Frank, Gumbel and Clayton copulas.

Note that the copula function can be time-varying through its parameters, which can be functions of past data.

Simulation algorithm:

At each simulation step do the following:

1. Fit univariate ARMA-GARCH models with marginal innovations distributions:
$$Z_i \sim F_i(z_i), i = \overline{1, N}$$
and estimate parameters of marginal distributions F_i .
2. Get empirical GARCH standardized residuals:
$$\hat{z}_{it} = \frac{r_{it} - \hat{\mu}_{it}}{\hat{\sigma}_{it}}, i = \overline{1, N}, t = \overline{1, n}$$
3. Transform Z_i into random variables that are distributed uniformly between zero and one, which removes individual information from the multivariate density F:
$$\hat{u}_{it} = F_i(\hat{z}_{it})$$
4. Fit copula function C to $\{\hat{u}_{1t}, \hat{u}_{2t}, \dots, \hat{u}_{Nt}, t = \overline{1, n}\}$ to get the parameters' estimates.
5. Simulate data from the fitted copula $u_1, u_2, \dots, u_N$.
6. Obtain simulated innovations by using inverse transformations
$$z_i = F_i^{-1}(u_i), i = \overline{1, N}$$
7. Simulate AR-GARCH processes with innovations z_i drawn from the copula.

DCC-GARCH

A related approach is to separate out correlation modeling from volatility modeling within a GARCH-type framework. The DCC GARCH has the flexibility of univariate GARCH but not the complexity of multivariate GARCH. These models are naturally estimated in two steps: DCC GARCH model firstly defines the individual conditional variances by using the univariate GARCH models, then the dependence between the individual series is specified via the covariance matrix.

Unlike the constant conditional correlation (CCC) the DCC model assumes that the correlation among assets varies over time which is in-line with the vast amount of empirical evidence supporting time-dependent correlation. Particularly, it is observed that correlation can be higher during periods of higher volatility associated with some market stress.

The conditional covariance matrix consists of two components that are estimated separately: time-dependent correlations R_t and the diagonal matrix of time varying volatilities D_t. Then, the covariance matrix is given by [3]:

$$H_t = D_t R_t D_t$$

where

$$D_t = diag\{\sigma_{1t}, \dots, \sigma_{nt}\}$$

the volatility of each asset σ_{it} follows a univariate GARCH process, R_t is a time-varying conditional correlation matrix of the following structure:

$$R_t = diag\{q_{11,t}^{-\frac{1}{2}}, \dots, q_{NN,t}^{-\frac{1}{2}}\} Q_t diag\{q_{11,t}^{-\frac{1}{2}}, \dots, q_{NN,t}^{-\frac{1}{2}}\}$$

$Q_t = \left(q_{ij,t}\right)$ is a symmetric positive definite covariance matrix which autoregressive dynamics follows the equation:

$$Q_t = (1 - A - B)\bar{Q} + A\varepsilon_{t-1}\varepsilon_{t-1}^T + BQ_{t-1}$$

$\bar{Q}$ is the $N \times N$ unconditional correlation matrix of u_t and A and B are non-negative scalar parameters satisfying $A + B < 1$ to ensure positive definiteness and stationarity, respectively. The elements of $\bar{Q}$ can be estimated or alternatively set to their empirical counterpart.

This model guarantees the positive definiteness of H_t if R_t is positive definite. The limiting assumption of the model is that the parameters A and B are assumed to be constants, which implies that the conditional correlations of all assets are driven by the same underlying dynamics.

Simulation algorithm for DCC AR(1)-GARCH(1,1):

Initialization

1. r_{i0} – last historical return's observation
2. $\mu_{i0} = \frac{\phi_{i0}}{1-\phi_{i1}}$ – unconditional mean
3. $\sigma_{i0}^2 = \frac{\omega_{i0}}{1-\alpha_{i0}-\beta_{i0}}$ – unconditional variance
4. $\bar{Q}$ – sample correlation matrix of standardized residuals from GARCH model calibration
5. Q_0– sample covariance matrix of standardized residuals from GARCH model calibration for the last M observations
6. ε_{i0} – last standardized residual from GARCH model calibration

Simulation

1. $\sigma_{it}^2 = \omega_i + \alpha_i(r_{it-1} - \mu_{it-1})^2 + \beta_i\sigma_{it-1}^2$
2. $\mu_{it} = \phi_{i0} + \phi_{i1}r_{it-1}$
3. $D_t = diag\{\sigma_{1t}, \dots, \sigma_{nt}\}$
4. $Q_t = (1 - A - B)\bar{Q} + A\varepsilon_{t-1}\varepsilon_{t-1}^T + BQ_{t-1}$
5. $R_t = diag\{q_{qq,t}^{-\frac{1}{2}}, \dots, q_{NN,t}^{-\frac{1}{2}}\} Q_t diag\{q_{qq,t}^{-\frac{1}{2}}, \dots, q_{NN,t}^{-\frac{1}{2}}\}$
6. $H_t = D_t R_t D_t$
7. $\mathrm{Z}_t \sim t(\nu)$
8. $a_t = H_t^{1/2}\mathrm{Z}_t$
9. $r_{it} = \mu_{it} + a_{it}$
10. $\varepsilon_{it} = \frac{a_{it}}{\sigma_{it}} = \frac{r_{it}-\mu_{it}}{\sigma_{it}}$

Multivariate Student-t GARCH

The innovations of the univariate GARCH process can be simulated from a multivariate distribution to capture the correlation structure of the asset returns. The multivariate normal and the multivariate Student-t distributions belong to the most widely used multivariate distributions for this purpose. In comparison to the sample from the multivariate normal distribution, the multivariate t distribution shows significantly heavier tails hence it is more suitable for modeling purposes.

Let assume that in one-dimensional GARCH models, N-dimensional random vector Z of innovations is drawn from the multivariate Student-t distribution with given parameters.

Let the N-dimensional random vector Y has multivariate normal distribution $N(0_{N\times 1}, \Sigma)$ and scalar random variable u has chi-squared distribution with ν degrees of freedom, Y and u are independent, Σ is $N \times N$ positive definite matrix. Then the N-dimensional random vector $Z = \sqrt{\frac{u}{\nu}}Y + \mu$ is said to be distributed as a multivariate t-distribution with parameters $\mu = [\mu_1, \dots, \mu_N]^T, \Sigma, \nu$ and has the probability density function

$$f_Z(z) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)(\nu\pi)^{N/2}\sqrt{\det\Sigma}}\left[1 + \frac{1}{\nu}(z-\mu)^T\Sigma^{-1}(z-\mu)\right]^{-(\nu+N)/2}, z \in \mathbb{R}^N$$

The covariance matrix of a multivariate Student's t random vector Z is well-defined only when $\nu > 2$ and it is

$$Var[Z] = \frac{\nu}{\nu-2}\Sigma, \nu > 2$$

To simulate random vector Z from the multivariate Student-t distribution the following steps should be made:

1) Generate $u \sim \chi^2_\nu$ and $Y \sim N(0, \Sigma)$ independently,
2) Compute $Z = \sqrt{\frac{u}{\nu}}Y + \mu$.

Applied to the GARCH model, univariate models are first calibrated independently, then during simulation, innovations are generated from a multivariate Student-t distribution with zero mean and a given correlation matrix and degrees of freedom parameter. The assumption underlying this approach is that the same degree of freedom parameter is assumed for all assets, however, as empirical observations show, the DOF parameter can differ between assets (see Table 1).

2 Numerical Experiments

Input Data

Numerical experiments were carried out on the group of : AAVE, COMP, ETH, LUNA, SNX, ANC, MIR. The calibration period contains data from 17/03/2021 to 08/12/2021. Figures below displays actual prices' and returns' paths (prices are scaled for illustration purposes).

We use 15-minute interval as the time duration to run simulations, however the interval can be extended or narrowed depending on the purpose of the analysis. Specifically, for longer-term analysis, it is necessary to use at least a daily interval. For intraday analysis, one can use more high-frequent data.

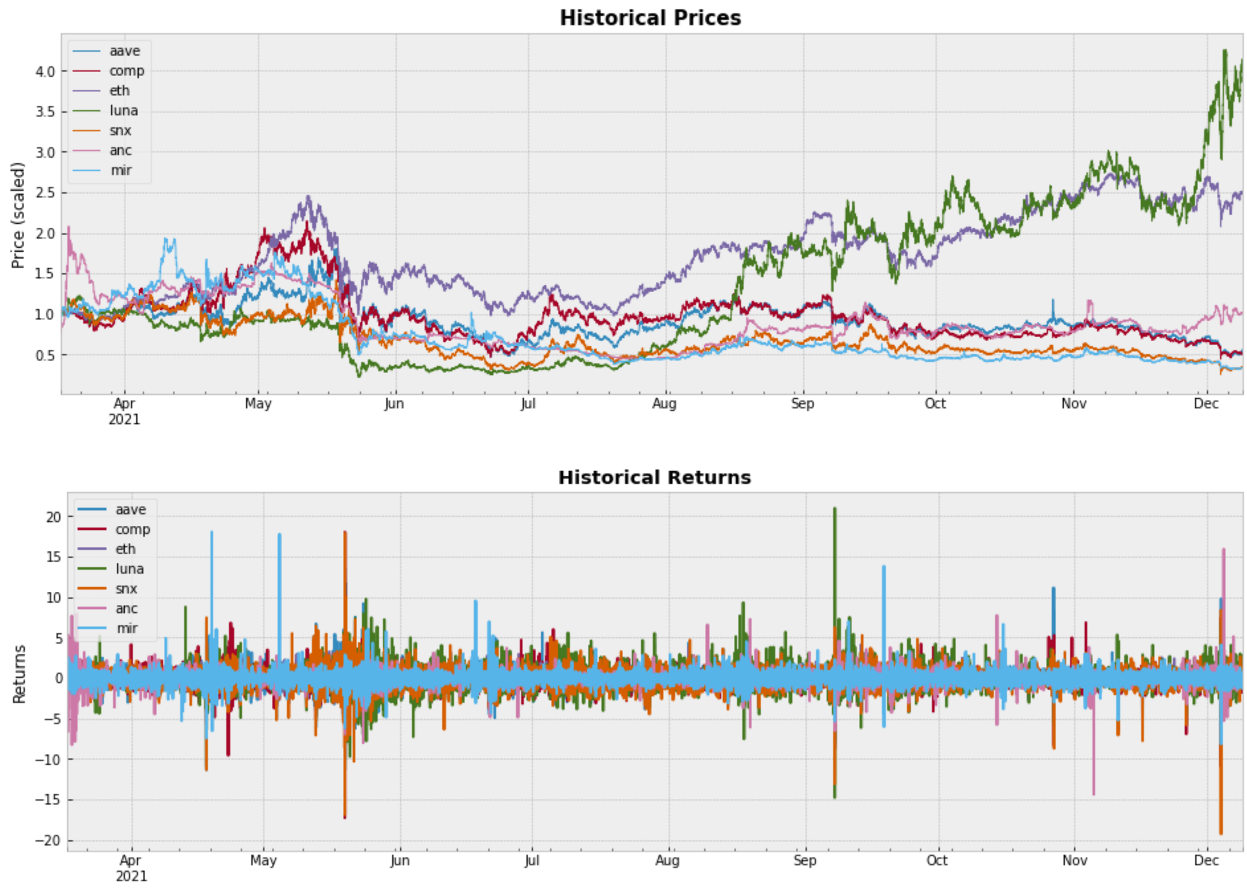

Figure 4. Historical 15-minute prices and corresponding simple returns

The following models have been implemented:

- Multivariate Brownian motion;
- Copula AR(1)-MGARCH(1,1);
- DCC AR(1)-MGARCH(1,1);
- Student-t AR(1)-MGARCH(1,1).

All MGARCH models used a two-step calibration procedure: first, one-dimensional GARCH models with Student-t residuals were calibrated for each asset independently by using the maximum likelihood method, then the model was estimated on the residuals of univariate GARCH models to take into account the correlation data structure. Calibrated models were used to simulate price trajectories 25 000 intervals into the future.

The applied approaches were evaluated according to the following criteria:

- difficulties of implementation and calibration;
- simulation speed 1000 trajectories 25000 intervals long;
- closeness of the correlation structure of simulated and real observations;
- closeness of the distribution laws of simulated and real observations.

Goodness-of-fit Tests

The following major points should be considered when comparing actual and simulated prices and assessing the quality of the model used for simulation:

1) the similarity of distribution;
2) the similarity of correlation structure.

Some formal tests are provided below. The below tests can be performed for each particular realization of simulated paths, then results can be aggregated e.g. by estimating the percentage of simulations when test results were satisfactory.

It should be noted that when performing these tests, it is assumed that all data are produced by the same distribution with constant characteristics (mean, variance, correlation etc.), however, the time-series can be non-stationary, i.e. represent a composition of several distributions.

The coincidence of the correlation structure is most often assessed through a comparison of the Pearson correlation coefficients which are sensitive only to a linear relationship between two variables (which may exist even if one is a nonlinear function of the other). However actual dependencies can be non-linear making the correlation coefficient and the related comparison not representative in that case.

Clearly, more advanced tests can be performed for quality analysis of the simulated data.

Distribution Test

To check whether two underlying one-dimensional probability distributions of actual and simulated returns differ, the Kolmogorov–Smirnov two-sample test (KS-test) or Lehmann-Rosenblatt test are usually used.

Let $F_{1,n}(x), F_{2,m}(x)$ are empirical distribution functions built on the independent samples of the size n and m respectively. To test the null hypothesis that both samples belong to the same distribution, the Kolmogorov-Smirnov statistics is calculated as the largest absolute difference between the two distribution functions across all x values:

$$D_{n,m} = \sup_x \left| F_{1,n}(x) - F_{2,m}(x) \right|$$

If the following inequality satisfies

$$\sqrt{\frac{nm}{n+m}} D_{n,m} > K_\alpha$$

where K_α is the α-quantile of the Kolmogorov-Smirnov distribution, the null hypothesis on the homogeneity of the distributions is declined.

Correlation Test

To check the similarity of the correlation structure, multiple pair tests can be made. Suppose r_1 and r_2 are two Pearson correlation coefficients calculated from the independent samples of size n and m respectively. Firstly, we should apply the Fisher transformation of the correlation coefficient r as follows:

$$r' = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right)$$

This transforms r which is between -1 and 1 into r', which ranges the whole real number line. It turns out that r' is approximately normal, with nearly constant variance that only depends on the number of data points and not on r. If z is defined as follows

$$z = \frac{r_1' - r_2'}{s}$$

where

$$s = \sqrt{\frac{1}{n-3} + \frac{1}{m-3}}$$

then $z \sim N(0,1)$. If $|z| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the quantile of the standard normal distribution, the null hypothesis on the equality of two correlation coefficients is rejected.

Note that more advanced tests can be used to test the hypothesis that the correlation matrices are equal. As noticed by some researchers the pairwise comparison of correlation matrix is not a scalable solution when there are multiple stocks involved in the portfolio. In the paper [2], a new fluctuation test for constant correlation matrix is proposed under a multivariate setting. Instead of the vector of successively calculated pairwise correlation coefficients, the largest eigenvalue of sample correlation matrix is used as a test statistics.

Simulation Results

In this section, we provide the simulation results of the considered approaches.

Brownian motion

The below is the plot of a single Brownian motion simulation of the prices with the correlation structure estimated from real data.

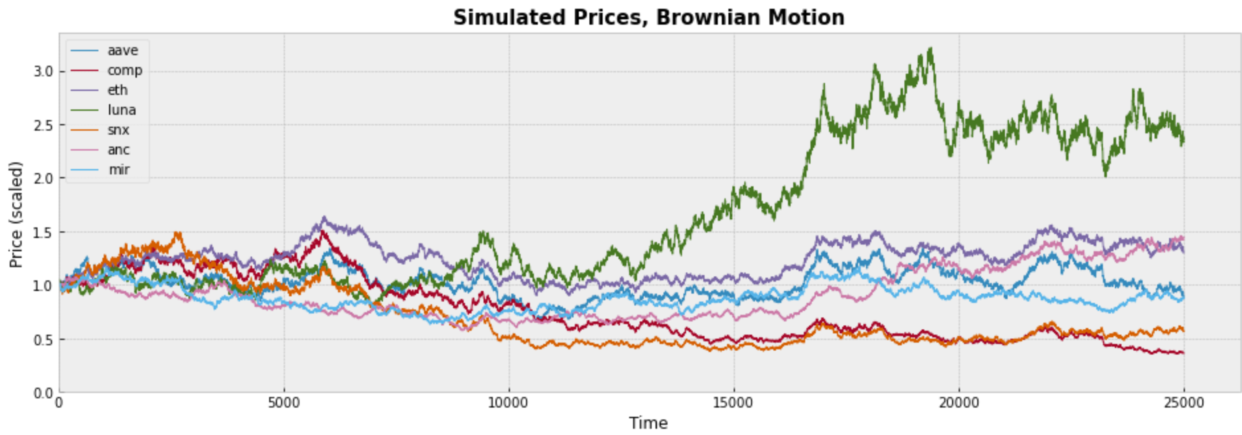

Figure 5. Simulated 15-minute prices, Brownian motion

Real and simulated asset returns are provided in the figures below.

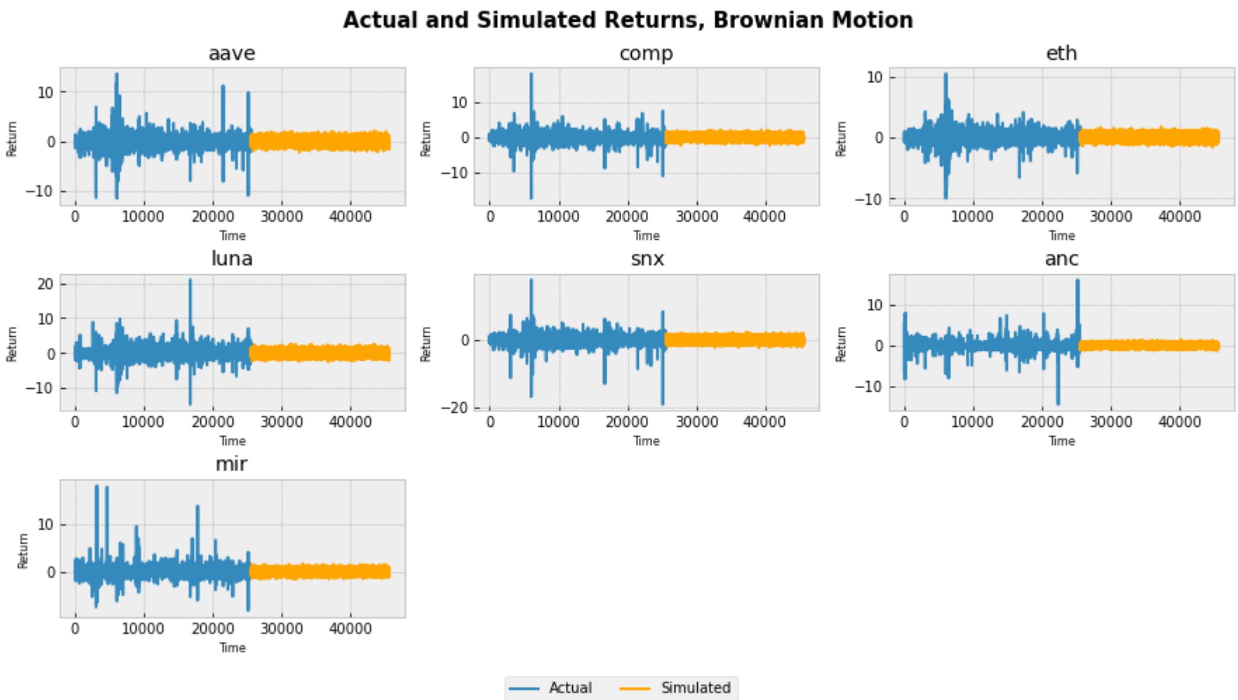

Figure 6. Actual and simulated 15-minute returns, Brownian motion

As can be seen from the above figures, the simulated returns do not repeat the patterns observed in real data, in particular, volatility does not change over time, there are no volatility clusters and sharp spikes.

The following are histograms of the distribution of real and simulated returns in logarithmic scale and corresponding p-values for the Kolmogorov-Smirnov test.

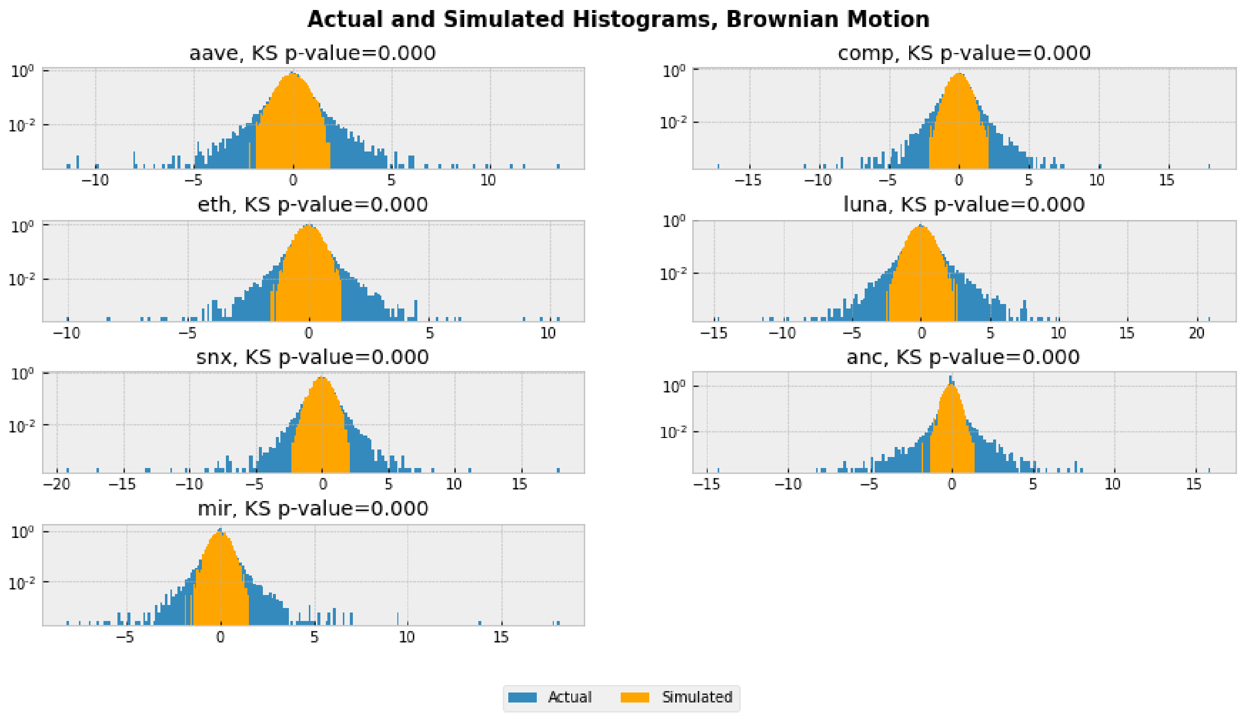

Figure 7. Actual and simulated histograms of the 15-minute returns, Brownian motion

It can be seen from the given histograms that the distributions of real and simulated data do not match, the heavy tails of the distribution are not taken into account, meaning that extreme returns in simulations are almost impossible.

Below are scatterplots for all pairs of assets, reflecting the real and simulated correlation structure, and p-values for the correlation equality test are also given.

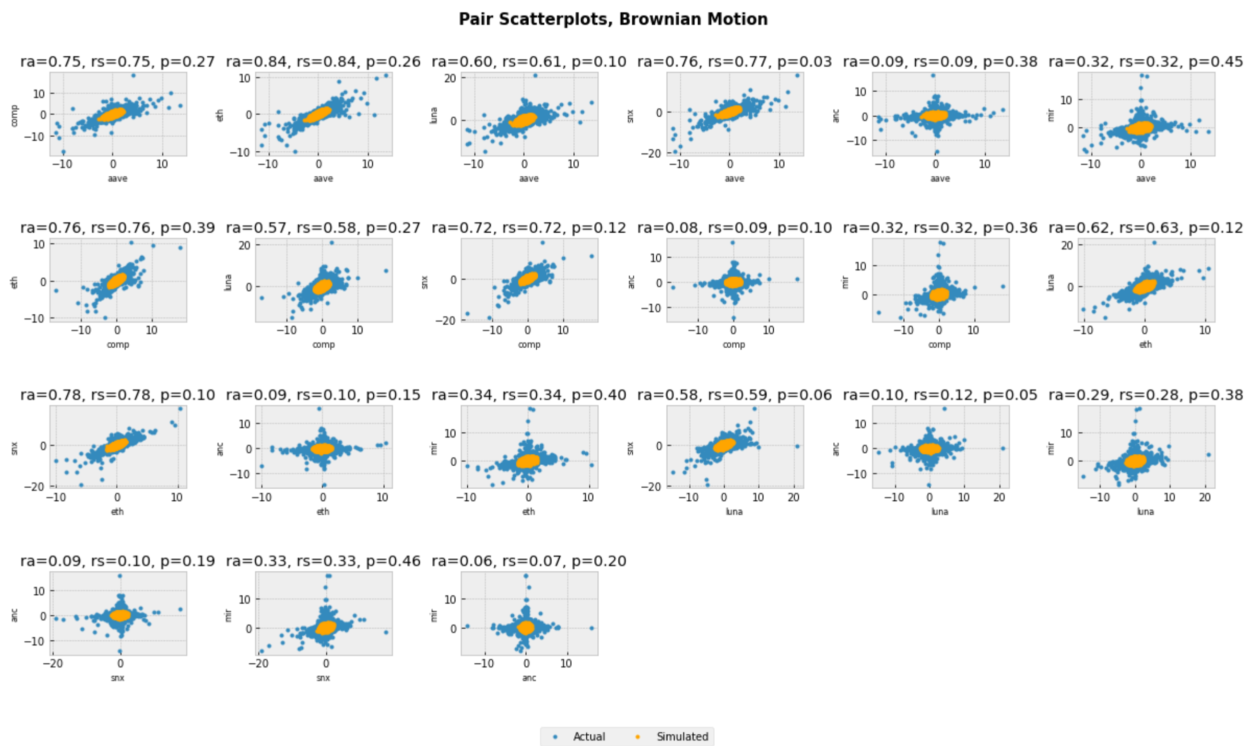

Figure 8. Actual and simulated pair scatterplots of the 15-munute returns, Brownian motion

It can be noticed that the pair Pearson correlations between the assets quite close for all pairs (all p-values are above the critical level of 5%). However, it can be seen that the real dependence is often far from linear and simulated data does not reflect this dependence. The linear correlation coefficient in this case is not an appropriate measure of dependence.

We conclude that despite the ease of implementation and speed of calculations, Brownian motion is a quite primitive tool for simulating price trajectories that does not reflect the features of real data. The use of such trajectories for risk assessment can lead to significant underestimations of realized losses.

Copula MGARCH

The below is the plot of a single Gaussian copula MGARCH model simulation of the prices with the correlation structure estimated from real data.

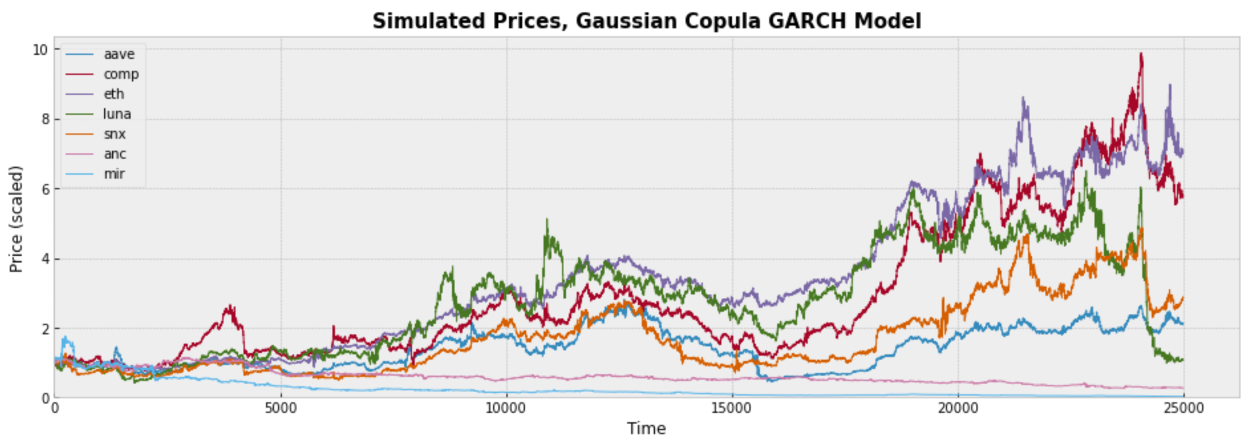

Figure 9. Simulated 15-minute prices, copula-MGARCH model

Real and simulated asset returns are provided in the figures below.

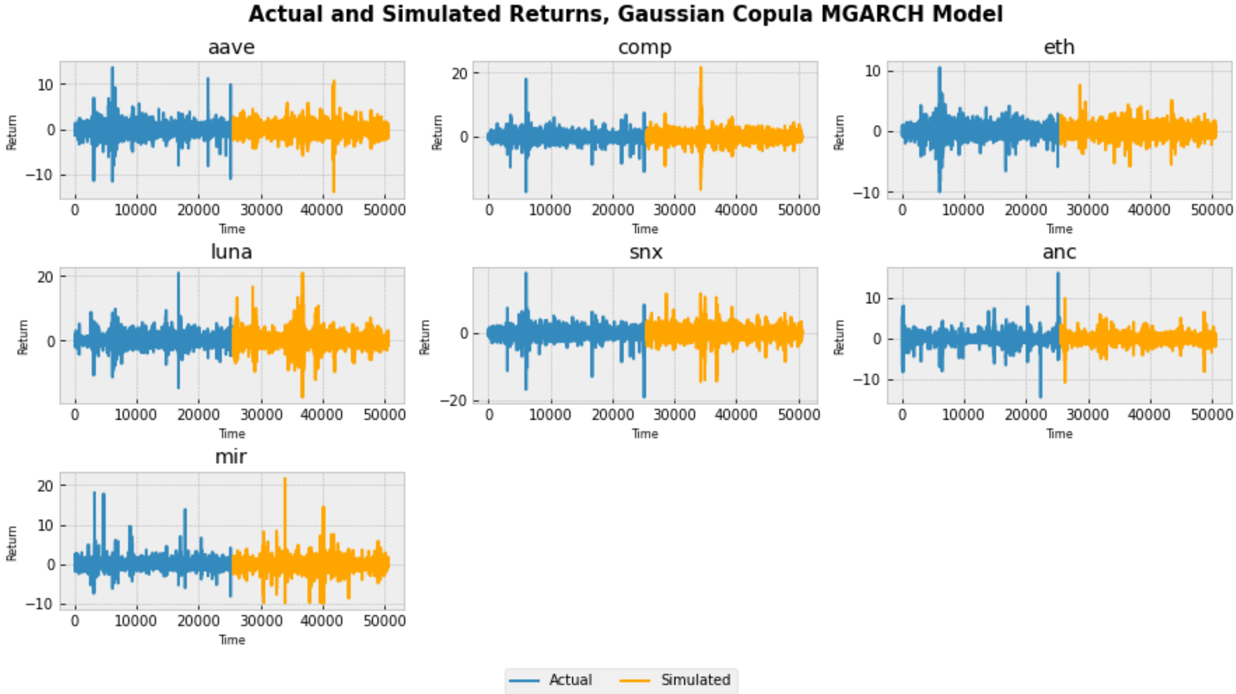

Figure 10. Actual and simulated 15-minute returns, copula-MGARCH model

It can be seen that the simulated prices look much more extreme compared to the Brownian motion model. The simulated returns in general reflect the historical behavior of real data including significant spikes, time-varying volatility and volatility clusters.

Histograms of real and simulated returns for each asset in logarithmic scale and corresponding p-values for the Kolmogorov-Smirnov test are provided below.

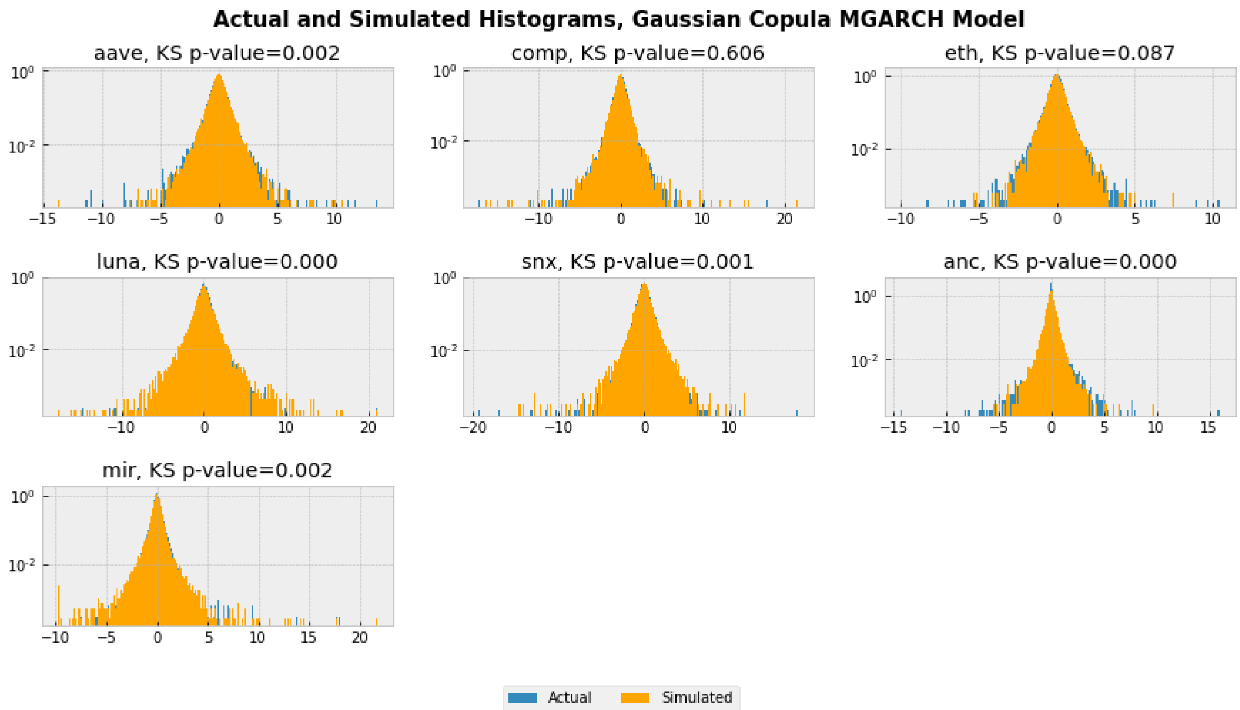

Figure 11. Actual and simulated histograms of the 15-minute returns, copula-MGARCH model

Visual inspection of the above histograms shows that the form of distribution (kurtosis and skewness) and heavy tails of distributions are captured quite well, however, formally according to the KS-test the distributions coincide at a significance level of 5% only for 2 assets out of 7. This means that in real data there can be some other features that are not taken into account in the classical GARCH model, in particular, some tail asymmetry is possible.

Below are scatterplots for all pairs of assets, reflecting the real and simulated correlation structure; p-values for the correlation equality test are also provided in the figures.

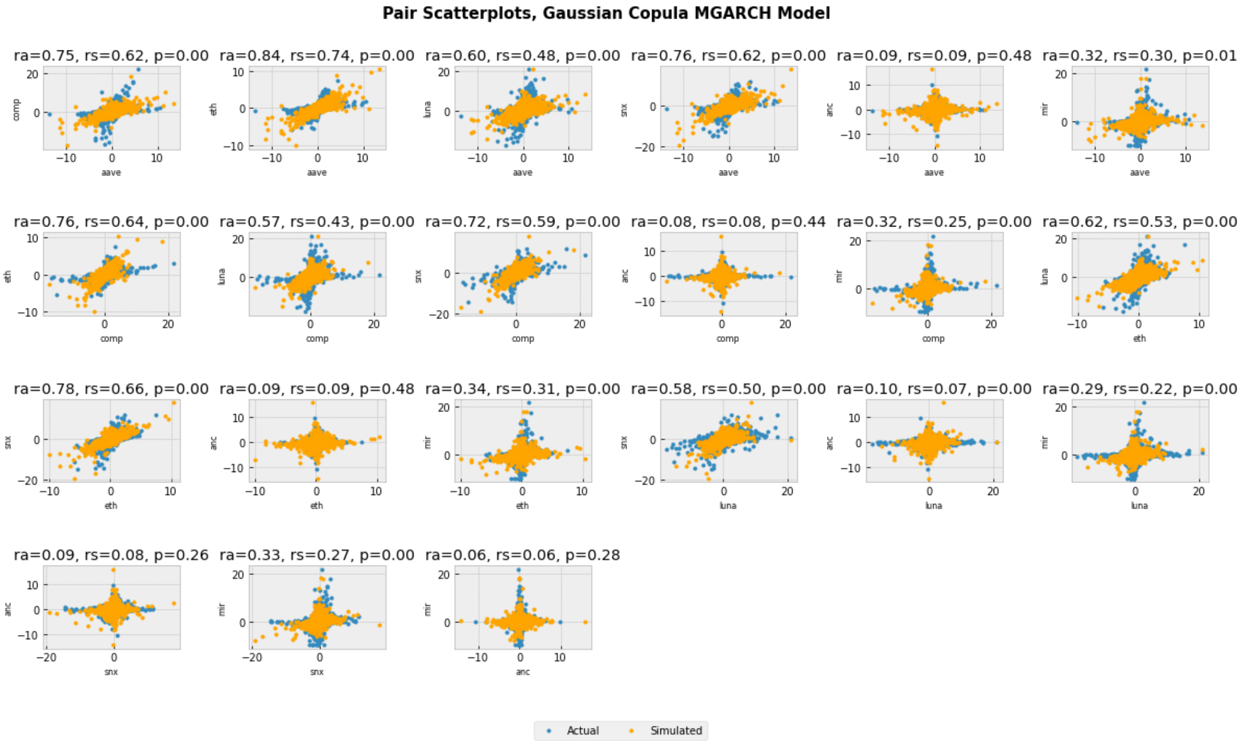

Figure 12. Actual and simulated pair scatterplots of the 15-munute returns, copula-MGARCH model

As seen the values of Pearson's pairwise correlation coefficients for simulated data are lower than the actual ones for most of pairs. However, it is worth mentioning that for some pairs the model managed to reproduce the non-linear nature of the dependence (see, e.g., 'anc-eth', 'anc-luna', 'mir-snx').

It can be concluded that the copula-MGARCH model is quite promising, since it has an acceptable quality and allows taking into account some important features of real data like volatility clusters, heavy tails and non-linear dependence between assets. Further extensions and improvements are also possible, in particular, other analytic copula functions or empirical copulas can be considered as well as copulas with time-varying parameters.

DCC-MGARCH

The below is the plot of a single simulation of the prices obtained using the DCC-MGARCH model.

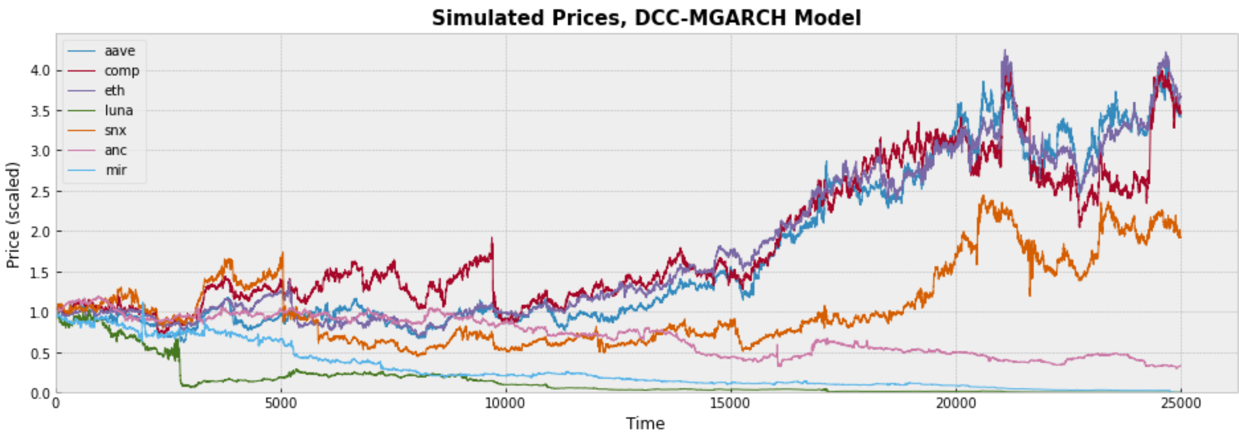

Figure 13. Simulated 15-minute prices, DCC-MGARCH model

Actual and simulated asset returns and corresponding histograms are provided in the figures below.

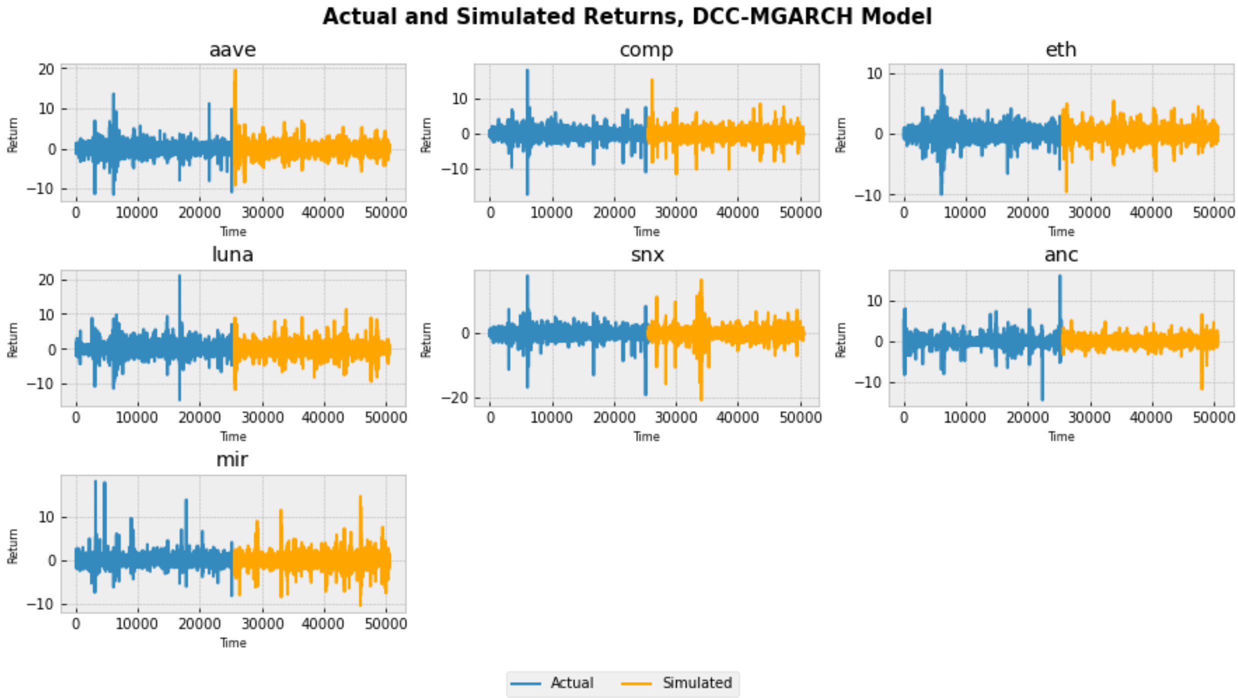

Figure 14. Actual and simulated 15-minute returns, DCC-MGARCH model

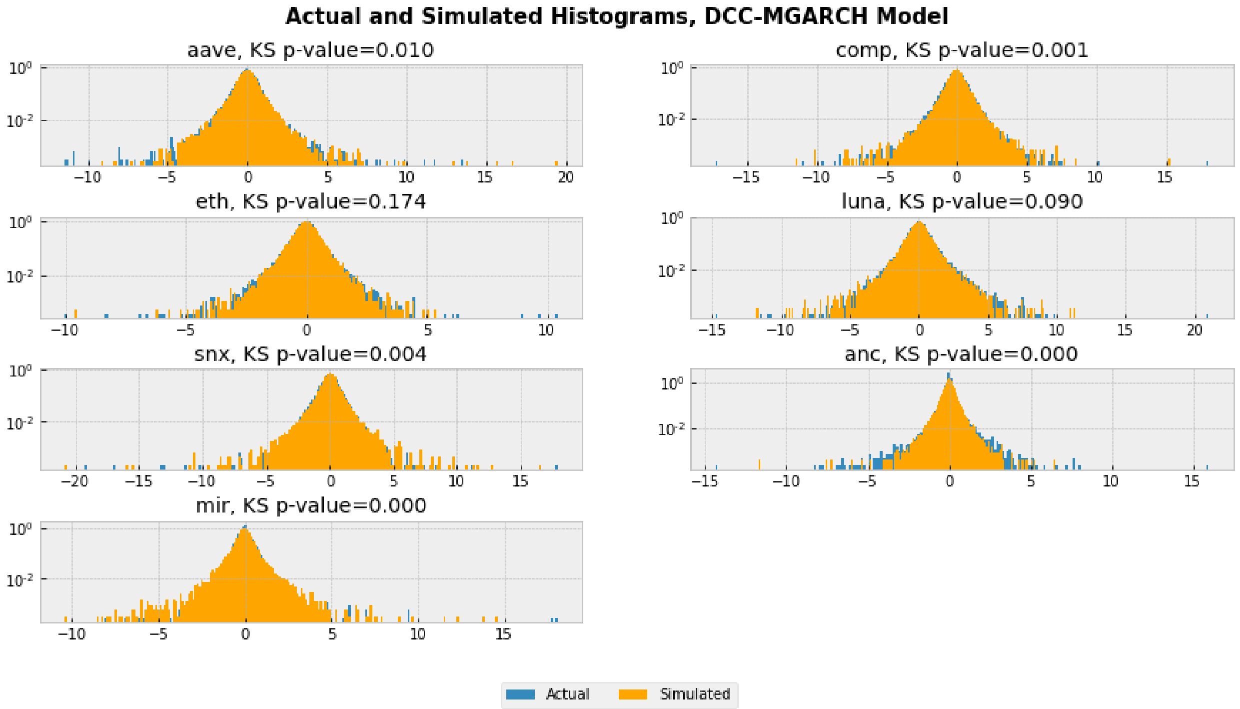

Figure 15. Actual and simulated histograms of the 15-minute returns, DCC-MGARCH model

As seen from the histograms, the hypothesis on the similarity of distributions is not rejected for 3 assets out of 7 (‘aave’, ‘eth’ and ‘luna’) at a significance level of 5%.

Below are scatterplots for all pairs of assets, reflecting the real and simulated correlation structure; p-values for the correlation equality test are also given.

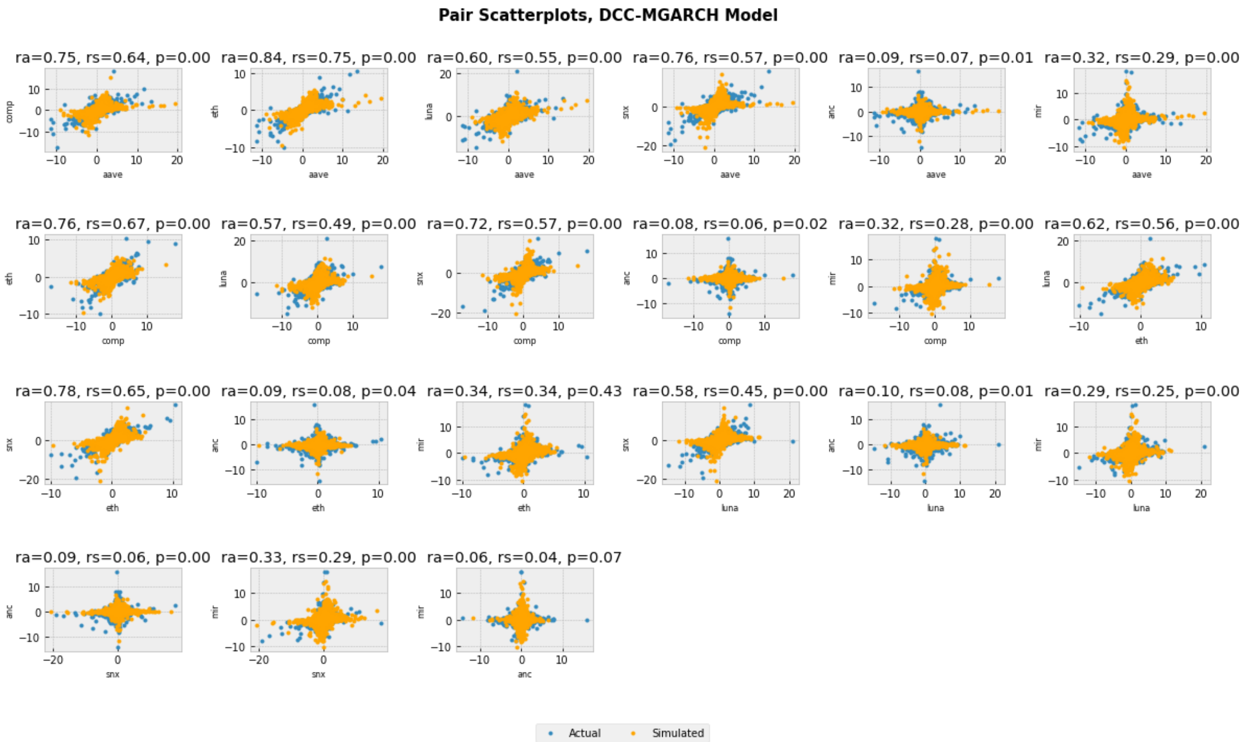

Figure 16. Actual and simulated pair scatterplots of the 15-munute returns, DCC-MGARCH model

As seen the simulated dependence is similar to the real one for some pairs of assets (e.g. ‘mir’-‘eth’, ‘anc’-‘aave’, ‘mir’-‘anc’). It is worth noting that the DCC model implies correlations that dynamically change over time, so tests for checking the coincidence of correlations over the entire sample of observations in this case are purely illustrative and cannot be correctly interpreted.

In general, we can say that the DCC-MGARCH model is quite promising and can be used for paths simulations. The advantage of the model is the usage of dynamic correlations confirmed by many empirical studies. Among the problems associated with the practical application of this model, one can note the difficulties with calibrating the model for dynamic correlations. A significant limitation is the mandatory positive definiteness of the correlation matrix, which creates difficulties in finding a solution for a fairly large number of assets. In future studies, other specifications of the DCC-model can be investigated.

Student-t MGARCH

The below is the plot of a single simulation of the prices obtained using the Student-t-MGARCH model.

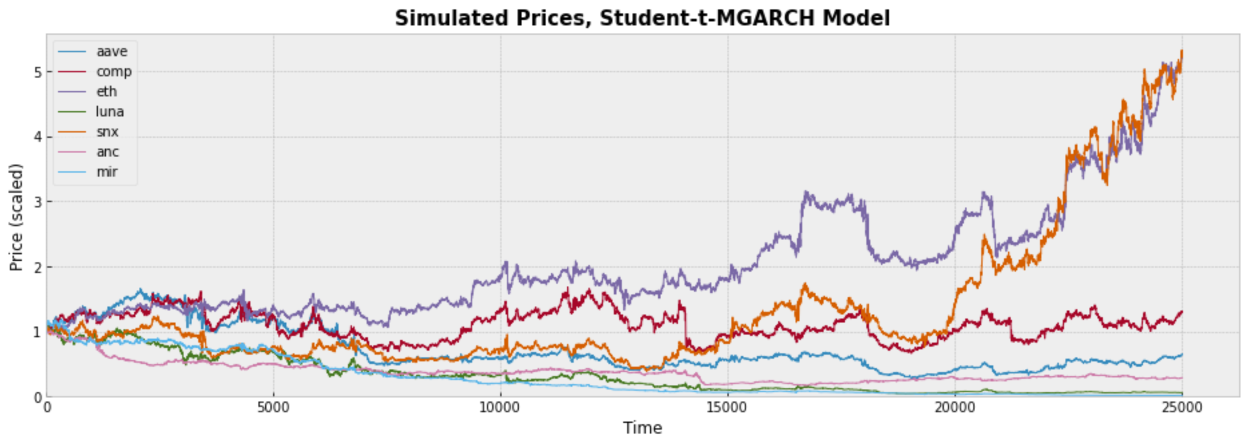

Figure 17. Simulated 15-minute prices, Student-t-MGARCH model

Real and simulated asset returns and their histograms are provided in the figures below.

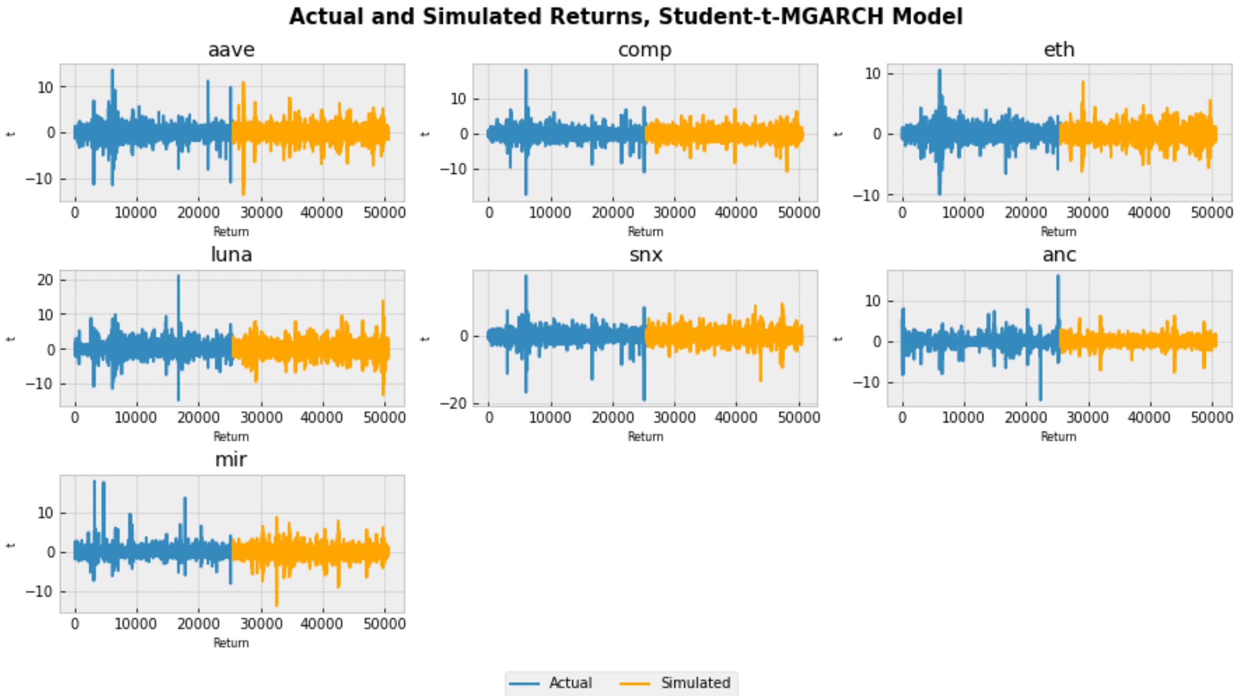

Figure 18. Actual and simulated 15-minute returns, Student-t-MGARCH model

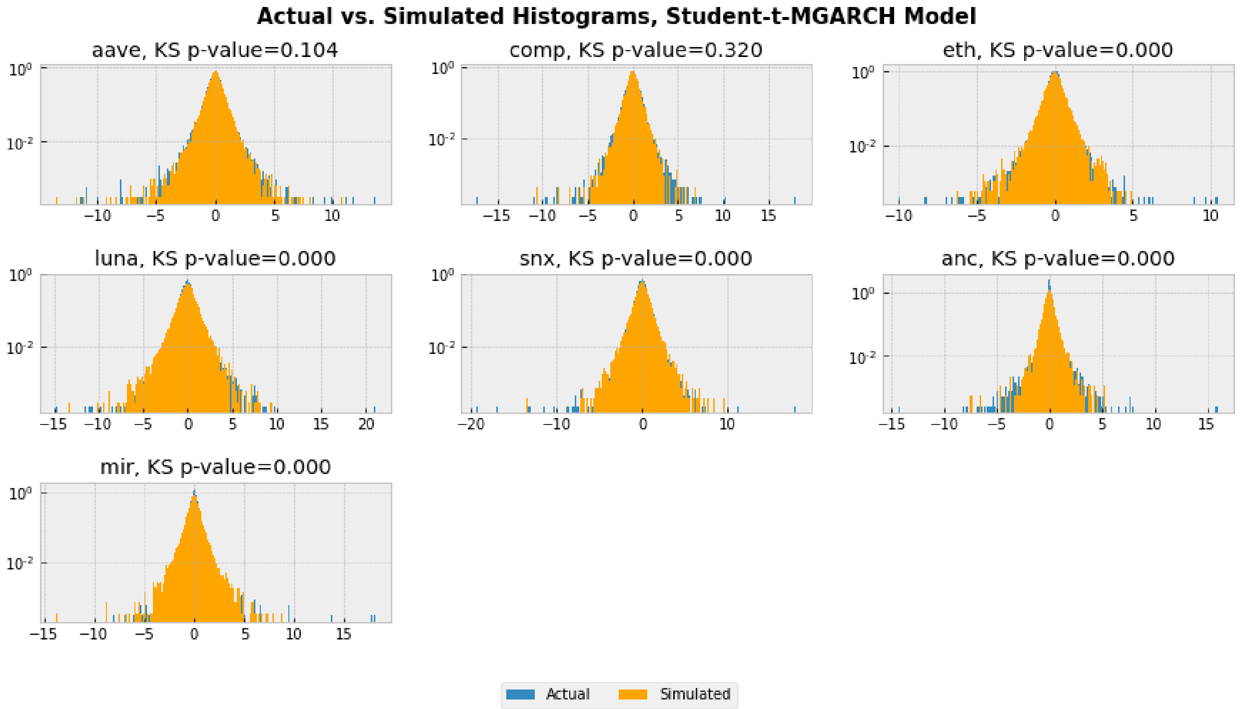

Figure 19. Actual and simulated histograms of the 15-minute returns, Student-t-MGARCH model

As seen from the histograms the hypothesis on the similarity of distributions is not rejected for 2 assets out of 7 ('aave' and 'comp') at a significance level of 5%.

Below are scatterplots for all pairs of assets, reflecting the real and simulated correlation structure; p-values for the correlation equality test are also given.

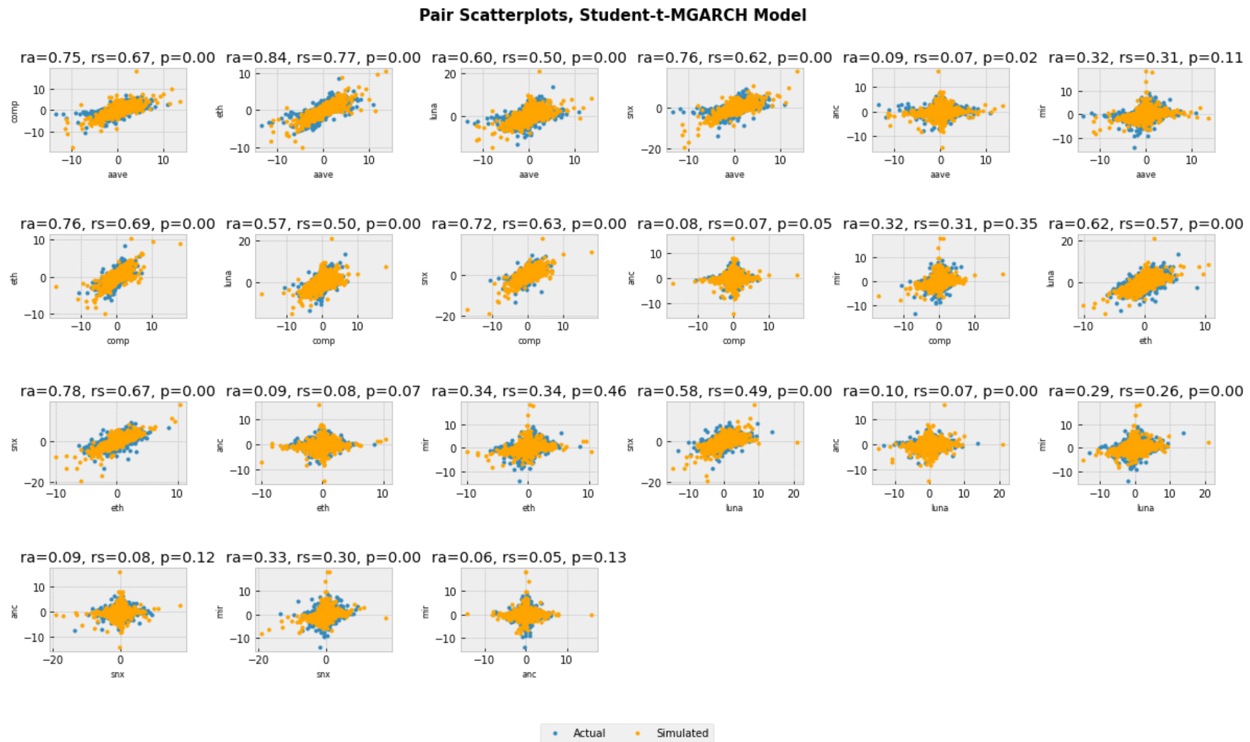

Figure 20. Actual and simulated pair scatterplots of the 15-munute returns, Student-t-MGARCH model

It can be seen from the visual inspection of the graphs that the dependence structure for simulated data is close to the real one for most of pairs, however, the correlation tests are formally failed in most cases.

Summarizing, the Student-t-GARCH model can be considered as appropriate, it is quite intuitive and is able to describe real data patterns and non-linear dependencies. Also the model is quite simple to implement and does not require complex extra-calibrations. Among the shortcomings, it is worth noting the underlying assumption on the same DOF parameter for all assets and a very slow speed of simulations, most of the time is spent on generation from the multivariate Student's distribution.

Conclusion and Future Research

In this study, different approaches were considered to simulate price trajectories of cryptocurrencies for the purpose of agent-based modelling including traditional geometric Brownian motion model and different specifications of the MGARCH model. The table below shows the main indicators for comparing the considered approaches.

#	Indicator	Brownian motion	Copula-MGARCH	DCC-MGARCH	Student-t MGARCH
1	Difficulty of implementation	Low	Middle	High	Low
2	Simulation speed (16Gb RAM) 25 000 intervals long	10 paths ~4 sec 1000 paths ~7 min	10 paths ~1 sec 1000 paths ~40 min	10 paths ~1 hour 1000 paths Huge	10 paths ~ 50 min 1000 paths Huge
3	Closeness of pair Pearson correlations	High	Low	Low	Low
4	Volatility clusters, fat tails	None	Yes	Yes	Yes
5	Dynamic correlations	None	None	Yes	None
6	Non-linear dependencies between assets	Low	High	High	High
7	Closeness of the distribution laws of simulated and real observations	Low	Middle	Middle	Middle

Table 2. Comparative analysis results of various approaches to price simulation

According to the results provided in the Table 2, Brownian motion model was found to be not appropriate as the simulated returns do not reflect real-data patterns. The MGARCH models considered (DCC-MGARCH, Copula-MGARCH, Student-t MGARCH) showed acceptable quality of the simulated data in terms of taking into account the features of real data, such as time-varying volatility, volatility clusters, heavy tails of returns distribution, and non-linear relationships between assets.

Among the considered MGARCH models, the model with multivariate Student's innovations is the easiest to implement, but on standard computing power it works quite slowly. The DCC model implies dynamic correlations between assets, but it is also quite slow and requires extra calibration, which can lead to problems in finding a solution for a large number of assets. In terms of computing resources and quality, the recommended model is the copula-MGARCH model.

Further research can be directed to the following aspects:

- consider alternative specifications of univariate GARCH models to make distributions for all assets match (e.g. GARCH-M to add a heteroskedasticity term into the mean equation; NGARCH to reflect a phenomenon called "leverage effect", signifying that negative returns increase future volatility by a larger amount than positive returns of the same magnitude; QGARCH, TGARCH to model asymmetry of positive and negative shocks etc.);

- consider alternative copula functions and/or copulas with time-varying parameters in the copula-MGARCH model;
- consider other specifications of MGARCH models, e.g. OGARCH, GOGARCH, GDC-MGARCH, etc.;
- perform code optimization to reject simulation time.

References

[1] Bauwens L., Laurent S., and Rombouts J. V. K. (2006) Multivariate GARCH models: A Survey (2006) // J. Appl. Econ. 21, P. 79–109.

[2] Bhattacharyya M. and Kasa R. (2018) A simple test for constant correlation matrix [arXiv:1809.07114v1]

[3] Danı´elsson J. (2011) The Theory and Practice of Forecasting Market Risk, with Implementation in R and Matlab. John Wiley & Sons Ltd., 296 P.

[4] Engle R. (2002) Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models // Journal of Business & Economic Statistics. Vol. 20, No. 3, P. 339–350.

[5] Gauntlet Research Report. Market Risk Assessment: An analysis of the financial risk to participants in the Aave V1 and V2 protocols. April 2021.

[6] Hansen, R. P. and Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1,1)? // Journal of Applied Econometrics 20(7), P. 873-889.