

# README file

The scripts in this repository are designed to evaluate the performance of the research detailed in *Quantile and Expectile Copula-Based Hidden Markov Regression Models for the Analysis of the Cryptocurrency Market* by Foroni, Merlo, and Petrella (2024). This evaluation includes a simulation study of copula-based quantile and expectile Hidden Markov Models (CQHMM and CEHMM) and an application of these models to a dataset of financial returns.

## Prerequisites

### Software Requirements

- R version 4.1.2 or higher
- RStudio version 4.4.0 or higher

### R Packages used (version in parentheses)

- MASS (7.3.60.2)
- mvtnorm (1.2.4)
- copula (1.1.3)
- skewt (1.0)
- foreach (1.5.2)
- doParallel (1.0.17)
- parallel (4.4.0)
- ald (1.3.1)
- quantreg (5.97)
- expectreg (0.52)
- mclust (6.1.1)
- cluster (2.1.6)
- markovchain (0.9.5)
- rqPen (4.1)
- lqmm (1.5.8)
- stats4 (4.4.0)
- sn (2.1.1)
- tidyverse (2.0.0)
- ggplot2 (3.5.1)
- PerformanceAnalytics (2.0.4)
- tseries (0.10-56)
- stats (4.4.0)
- reshape2 (1.4.4)
- scales (1.3.0)
- ggpubr (0.6.0)
- dotwhisker (0.8.2)
- patchwork (1.2.0)
- plotly (4.10.4)

## Simulation\_1Y.R

This script evaluates the performance of the CQHMM and CEHMM models generating data from a bivariate two-states HMM using the following data generating process for  $t = 1, \dots, T$ ,

$$\mathbf{Y}_t = \mathbf{X}_t \boldsymbol{\beta}_k + \boldsymbol{\epsilon}_{t,k}, \quad S_t = k$$

where  $\mathbf{X}_t = (1, X_t)'$  and the true values of the regression parameters are

$$\beta_1 = \begin{pmatrix} -2 & 3 \\ 1 & -2 \end{pmatrix} \quad \text{and} \quad \beta_2 = \begin{pmatrix} 3 & -2 \\ -2 & 1 \end{pmatrix}.$$

It is possible to choose between two different scenarios for the transition probability matrix, scenario 1:

$$\mathbf{\Pi} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix},$$

scenario 2:

$$\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{pmatrix}.$$

### Running the Script

1. Open the `Simulation_1Y.R` script in RStudio.
2. At line 26, select the type of model (`expectile` or `quantile` for CEHMM and CQHMM respectively) to be used.
3. At line 37, set the number of cores to be used in the parallel computation.
4. Define simulation setting by choosing the number of observations (`n`), the errors distribution (`dist`), the transition probability matrix (`gamma_setting`), the number of states (`k`), the number of restarts (`R`), the number of simulations (`MM`), the vector of quantiles or expectiles considered (`tauvec`) and the copula to be used (`wcop`).

### Simulation\_4Y.R

This script evaluates the performance of the CQHMM and CEHMM models considering a four-dimensional response variable ( $d = 4$ ), one sample size ( $T = 1000$ ) and three explanatory variables,  $X_t^{(1)}$ ,  $X_t^{(2)}$  and  $X_t^{(3)}$  sampled from independent standard Normal distributions. Observations are drawn from a three-state HMM using the following data generating process for  $t = 1, \dots, T$ ,

$$\mathbf{Y}_t = \mathbf{X}_t \beta_k + \epsilon_{t,k}, \quad S_t = k$$

where  $\mathbf{X}_t = (1, X_t^{(1)}, X_t^{(2)}, X_t^{(3)})'$  and the true values of the state-specific intercepts are drawn from Uniform distributions defined by:  $\mathcal{U}(-3, -1)$ ,  $\mathcal{U}(-1, 2)$ , and  $\mathcal{U}(3, 5)$ , while slope parameters are generated from the following uniform distribution for each state:  $\mathcal{U}(-2, 2)$ .

We consider the following transition probability matrix:

$$\mathbf{\Pi} = \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.05 & 0.9 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{pmatrix}.$$

### Running the Script

1. Open the `Simulation_5Y.R` script in RStudio.
2. At line 26, select the type of model (`expectile` or `quantile` for CEHMM and CQHMM respectively) to be used.
3. Define simulation setting by choosing the number of observations (`n`), the errors distribution (`dist`), the number of states (`k`), the number of restarts (`R`), the number of regressors (`nregs`), the number of dependent variables (`multi_dim`), the number of simulations (`MM`), the vector of quantiles or expectiles considered (`tauvec`) and the copula to be used (`wcop`).
4. At line 192, set the number of cores to be used in the parallel computation.

## MainFunctions\_\_cquant.R

This script contains the main functions for the CQHMM model. The functions are:

- `expreg.hsmm.multi`: starting from the true values of the beta parameters, this function generates regressors and dependent data from a multivariate Gaussian.
- `expreg.hsmm.multi.skewt`: starting from the true values of the beta parameters, this function generates regressors and dependent data from a t Student and skew-t distributions.
- `hsmm.multi.real`: this function generates data from a Gaussian or t copula with Asymmetric Laplace marginals.
- `em.hmm.cqereg`: this function estimates the parameters of the CQHMM model using the EM algorithm.

## MainFunctions\_\_cexp.R

This script contains the main functions for the CEHMM model. The functions are:

- `expreg.hsmm.multi`: starting from the true values of the beta parameters, this function generates regressors and dependent data from a multivariate Gaussian.
  - `expreg.hsmm.multi.skewt`: starting from the true values of the beta parameters, this function generates regressors and dependent data from a t Student and skew-t distributions.
  - `hsmm.multi.real`: this function generates data from a Gaussian or t copula with Asymmetric Normal marginals.
  - `em.hmm.cqereg`: this function estimates the parameters of the CEHMM model using the EM algorithm.
- 

## RealData.R

This script contains the code to estimate the CQHMM and CEHMM models on the cryptocurrency dataset. The script reads the log-returns, estimates the models, and produces the results.

### Dataset Description

The dataset is included as `returns.RData` in the `data` folder of this repository. The variable `ret.df` is a dataframe with 1348 observations and 10 variables containing daily returns of the followings: Bitcoin (BTC), Ethereum (ETH), Litecoin (LTC), Ripple (XRP), Bitcoin Cash (BCH), S&P 500 index (S&P500), S&P US Treasury Bond index (SPUSBT), US Dollar Index (USDIX), WTI Crude Oil (WTI), and Gold.

The dataset spans from July 25, 2017, to December 19, 2022, providing a comprehensive overview of the interactions between these markets over a significant period.

### Running the Script

1. Open the `RealData.R` script in RStudio.
2. At line 32, upload the dataset `returns.RData` from the `data` folder.
3. At line 36, select the type of model (`expectile` or `quantile` for CEHMM and CQHMM respectively) to be used.
4. At line 48, set the number of cores to be used in the parallel computation.
5. From line 62 define the regressors, the dependent variable, the number of states, the number of restarts, the vector of quantiles or expectiles considered, and the copula to be used.

## Fig&Tab\_\_out.R

This script contains the code to reproduce Table S14, Table S15, Figure S3, Figure S2 and Figure 1 of the paper. The script reads the RData files `returns.RData` and `price_xts.RData` from the `data` folder.

## Script Description

The script `Figures_out.R` generates the followings:

- Table S14: Descriptive statistics.
- Table S15: Empirical correlation matrix.
- Figure S3: QQ plots of standardized residuals.
- Figure S2: Cryptocurrencies daily normalized prices and log return series.
- Figure 1: Time series of returns for the five cryptocurrencies colored according to the two-state fitted models.