Intraday Volatility for High Frequency Datausing MC-GARCH Model

- by Yiwei Ang (8th July 2020)
- In this document, we introduce:
  - 1. Homoskedascity vs Heteroskedascity
  - 2. GARCH Model Introduction
  - 3. Extensions of GARCH
  - 4. **Introduction of MC-GARCH (Engle & Sokalska 2012)**
  - 5. Estimation Results and Analysis of MC-GARCH
    - a. Data Description
    - b. Heteroskedasticity and Normality Tests of the Return Series
    - c. Identifying the Conditional Mean Equation
    - d. Model Checking for the Mean Equation
    - e. Estimation of Daily Variance (h_t) Forecast
    - f. Fitting Performance of MC-GARCH
    - g. Intraday VaR Forecast
      - Kupiec's Test

1. Homoskedascity vs Heteroskedascity

Homoskedascity

- **Homoskedascity** refers to a condition in which the **variance of the residual**, or error term, in a regression model is **constant**.
- The error term does not vary much as the value of the predictor variable changes.

Heteroskedascity

- **Heteroskedasticity** describes the **irregular pattern of variation of an error term**, or variable, in a statistical model.
- Essentially, where there is heteroskedasticity, observations do not conform to a linear pattern. Instead, they tend to cluster.
- In financial asset, the volatility seems to vary during certain periods of time and depend on past variance, hence they incur this characteristic.

Conditional Heteroskedascity (Time-Varying Volatility)

- As a matter of fact, the financial market reacts in the presence of many factos:
  - Interest Rates
  - Politically disorder
  - Economic events such as non farm payrol
- Statistically, $Var[X_t|X_{t-1}, X_{t-2}...]$ to be non costant over time.

2. GARCH Model Introduction

ARCH(q)

- To model a time series using an ARCH process, let ϵ_t denote the error terms (return residuals, with respect to a mean process), i.e. the series terms.
- These ϵ_t are split into a stochastic piece z_t and a time-dependent standard deviation σ_t characterizing the typical size of the terms so that

$$\epsilon_t = \sigma_t z_t$$

- The random variable z_t is a strong white noise process.
- The series σ_t^2 is modeled by

$$\sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \sigma_t^2 = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$$

,

- where $\alpha_0 > 0$ $\alpha_0 > 0$ and $\alpha_i \geq 0,\ i > 0 \alpha_i \geq 0,\ i > 0.$

GARCH(p, q)

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \cdots + \alpha_q \epsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2$$

- Hence, GARCH(p, q) is equivalent to:

$$GARCH(p, q) = ARCH(q) + \sum_{k=1}^{p} \beta_i \sigma_{t-i}^2$$

- GARCH(1, 1) is mathematically:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

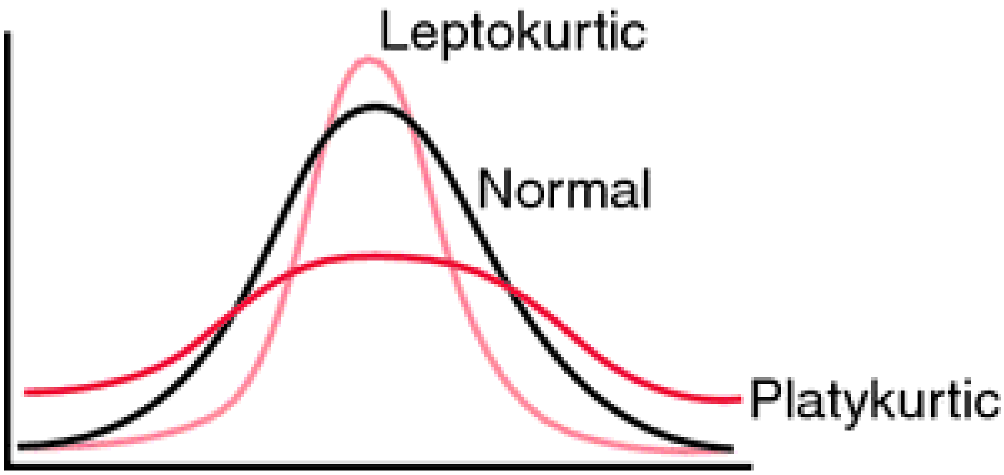
- and consists of:
  - 1 lagged term of conditional variance σ_{t-1}^2
  - 1 lagged term of square error u^2
- The **generalized autoregressive conditional heteroskedasticity (GARCH) process** is an econometric term developed in 1982 by Robert F. Engle, an economist and 2003 winner of the Nobel Memorial Prize for Economics.
- GARCH is preferred by financial modeling professionals because it provides a more **real-world context** than other forms when trying to predict the prices and rates of financial instruments.

Purpose

- **Estimate volatility in financial markets**, especially volatility of returns for stocks, bonds and market indices.
- **Determine pricing** and judge which assets will potentially provide **higher returns**.
- **Forecast the returns of current investments** to help in their asset allocation, hedging, risk management and portfolio optimization decisions.

Characteristics

- **Volatility Clustering**
  - Coefficient of θ_1 is usually high, around 0.9, for many weekly/daily financial time series.
  - Indicates large/small value of h_{t-1} will be followed by large/small value of h_t.
- **Fat Tails (Leptokurtic)**



  - High frequency financial data usually have its **return** fatter tails than normal distribution.
  - Indicates **small change is more frequent** than normal distribution.
- **Mean Reversion**
  - Long run variance is

$$\frac{b_0}{1-\theta_1-b_1}$$

General Process

- Estimate a best-fitting autoregressive model.
- Compute autocorrelations of the error term.
- Test for significance.

3. Extensions of GARCH

- **Threshold GARCH Model (T-GARCH)**
- **Exponential GARCH Model (E-GARCH) - Nelson(1991)**
  - Capture the leverage effect of shocks(policies, informations, news, events)
  - Good(Bad) news events results in assets price being less(more) volatile.

$$\sigma_t = \phi + \sum_{k=1}^{p} \theta_k \log(\sigma_{t-k}) + \sum_{k=1}^{q} \lambda_i \frac{u_{t-i}}{\sqrt{\sigma_{t-i}}} + \sum_{k=1}^{q} \eta_i | \frac{u_{t-i}}{\sigma_{t-i}} |$$

Flaws

- It has been developed mostly for **low frequency** of financial data.
- Has been proved to be inappropriate for **high frequency** which contains **intraday seasonality**.

4. Introduction of MC-GARCH (Engle & Sokalska 2012)

- https://www.researchgate.net/publication/330603428_Risk_Model_Validation_An_Intraday_VaR_and_ES_Ap (https://www.researchgate.net/publication/330603428_Risk_Model_Validation_An_Intraday_VaR_and_ES_Ap
- Assumes a **decomposition** of volatilities into **multiplicative components**.
- This paper developed a model for forecasting the volatility of intraday exchange rates using time series data.
- The MC-GARCH assumes the conditional variance to be the multiplicative product of daily volatility forecast, seasonal or diurnal volatility, and stochastic intraday volatility.
- The following carries out:

- The **daily volatility forecast** using **asymmetric Exponential GARCH (EGARCH(1,1)) model**.
- The **modelling and forecasting performance** of the MC-GARCH model.
- The forecasting and validation of **intraday Value-at-Risk (VaR)**.

- The results show that the MC-GARCH model is suitable to model and forecast the volatility and VaR of intraday EUR/USD exchange rates.

Model Specification

a) Daily Variance Component h_t

- Daily variance is modelled by the sum of :
  - Conditional Mean - ARIMA (p, d, q) and
  - Conditional Variance - GARCH/EGARCH(1, 1)

ARIMA(p, d, q)

$$\left(1-\sum_{i=1}^{p}\varphi_i L^i\right)(1-L)^d X_t=\delta+\left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t.$$

- where:
  - p is the number of autoregressive terms,
  - d is the number of nonseasonal differences needed for stationarity, and
  - q is the number of lagged forecast errors in the prediction equation

GARCH(1, 1)

- Specified by the following:
- **Daily Log Return Function**

$$r_t = m_t + \varepsilon t$$

where:

- m_t represents the conditional mean process made up of both autoregressive (AR) and moving averages (MA) terms
- r_t represents the daily log returns.
- ε_t is the error term which can be decomposed as $\varepsilon_t = h_t zt$
- and **Variance Equation**

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

where:

- h_t represents the volatility process to be estimated. The innovation term, zt are i.i.d. variables.
- To satisfy stationary, we need:
  - $\omega > 0,$
  - $\alpha_1 > 0,$
  - $\beta_1 > 0$
  - and $\alpha_1 + \beta_1 < 1$

EGARCH(1, 1)

- This is to capture the asymmetric effects between positive and negative asset returns and models the logarithm of the conditional variance h_t.

$$ln(h_t) = \omega + \frac{\alpha_1 \varepsilon_{t-1} + \gamma_1 |\varepsilon_{t-1}|}{h_{t-1}} + \beta_1 ln(h_{t-1})$$

b) MC-GARCH(1, 1) Model h_t

$$R_{t,i} = \sqrt{h_t s_i q_{t,i}} \varepsilon_{t,i}$$

- $R_{t,i}$ is the intraday return process.
- h_t is the daily variance component.
- s_i is the seasonal or diurnal variance component in each intraday period.
- $q_{t,i}$ is the intraday variance component
- $\epsilon_{t,i}$ is an error term (also called the standardized innovation) which follows a certain specified distribution, usually normal.

c) Steps of MC-GARCH(1, 1) Modelling:

ii. Daily volatility component h_t

- Based on Andersen and Bollerslev (1997), the study employs GARCH and EGARCH to forecast the daily variance component ht
- The choice of the model is based on the best-performing one among the GARCH and EGARCH models under five error distributions, which are:
  - Normal distribution,
  - Student's-t distribution,
  - Skewed Student's-t distribution,
  - Generalised Error Distribution (GED)
  - Johnson SU (JSU) distribution

ii. Diurnal volatility component s_i

- It is **intraday seasonality** component to capture and correctly distribute daily variance into intraday's.
- This is estimated as the variance of intraday returns in each **regularly spaced intraday time period** as represented below:

$$\frac{R_{t,i}^2}{h_t} = s_i q_{t,i} \varepsilon_{t,i}^2$$

$$s_i = \frac{1}{T} \sum_{t=1}^{T} \frac{R_{t,i}^2}{h_t}$$

- By using **daily and diurnal variance**, the returns are **normalized** by:

$$\begin{aligned} z_{t,i} &= \frac{R_{t,i}}{\sqrt{h_i s_i}} \\ &= \sqrt{q_{t,i}} \varepsilon_{t,i} \end{aligned}$$

iii. Stochastic intraday variance component $q_{t,i}$ - GARCH Process

$$q_{t,i} = \omega^* + \alpha_1^* \left(\frac{R_{t,i-1}}{\sqrt{h_t s_{i-1}}}\right)^2 + \beta_1^* q_{t,i-1}$$

- where:
  - $\omega^* > 0$,

- $\alpha_1^* \geq 0,$
- $\beta_1^* \geq 0,$

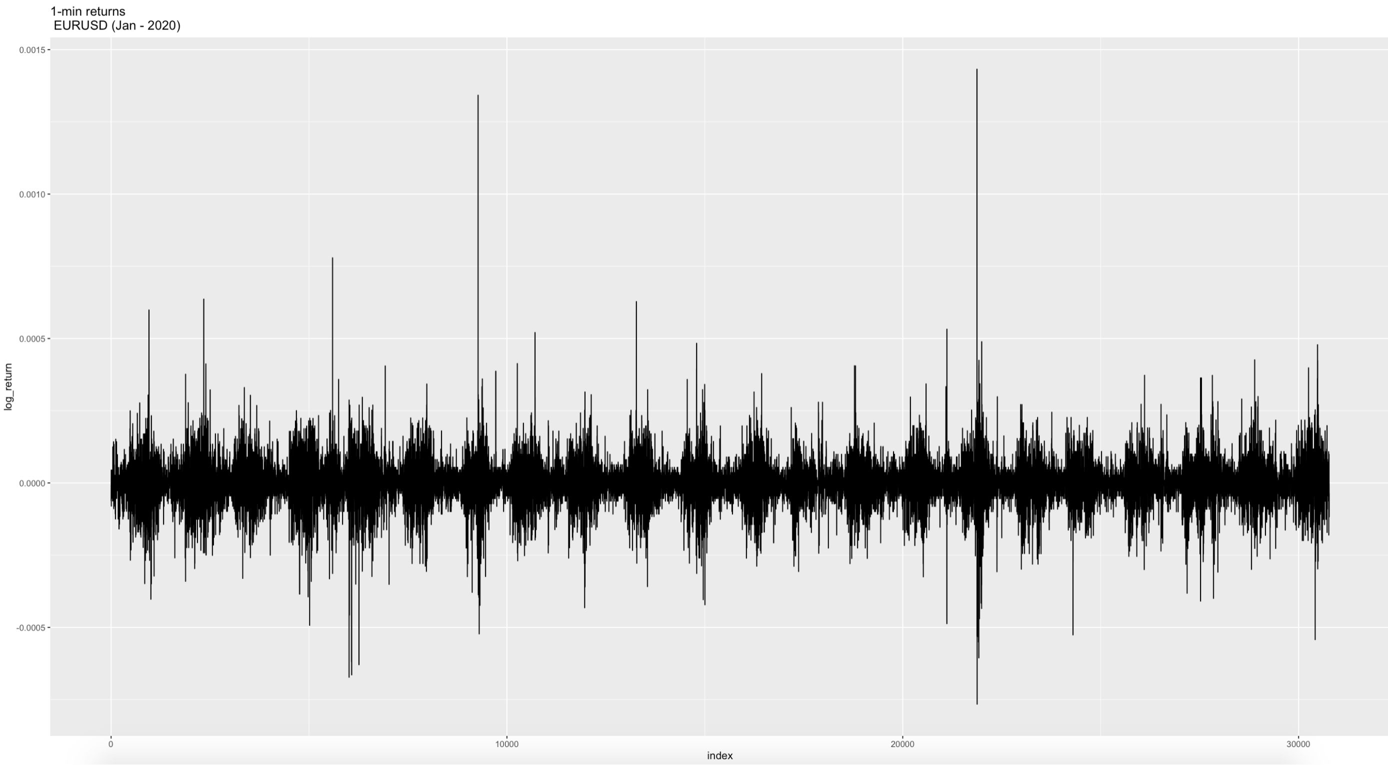
5. Estimation Results and Analysis

5.1 Data Description

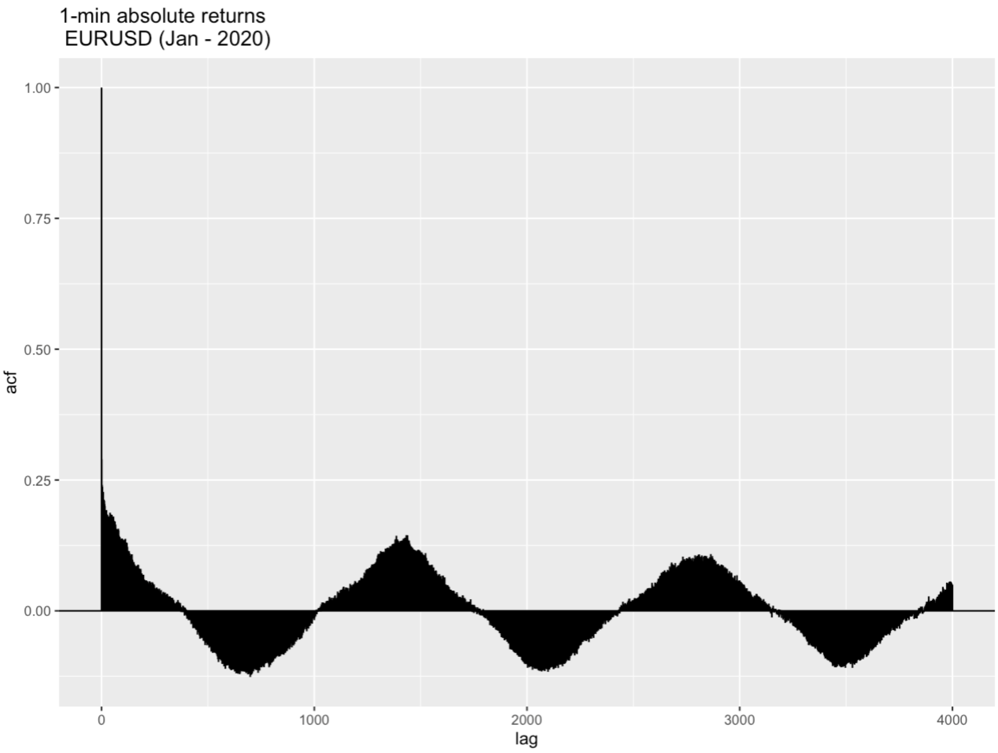
- The symbol to be analyzed in this documentation is EURUSD.
- For intraday 1 min data, we use Oct 2019 which consists of 28290 observations and equivalently 21 days of intraday logarithmic returns.
- For daily data, we need a longer period, 10 years from Oct 2009 to Oct 2019.
- The log-returns are calculated based on the following, to transform a financial time series into a stationary series:

$$r_t = ln(\frac{P_t}{P_{t-1}})$$

Log- Return of 1 min EURUSD

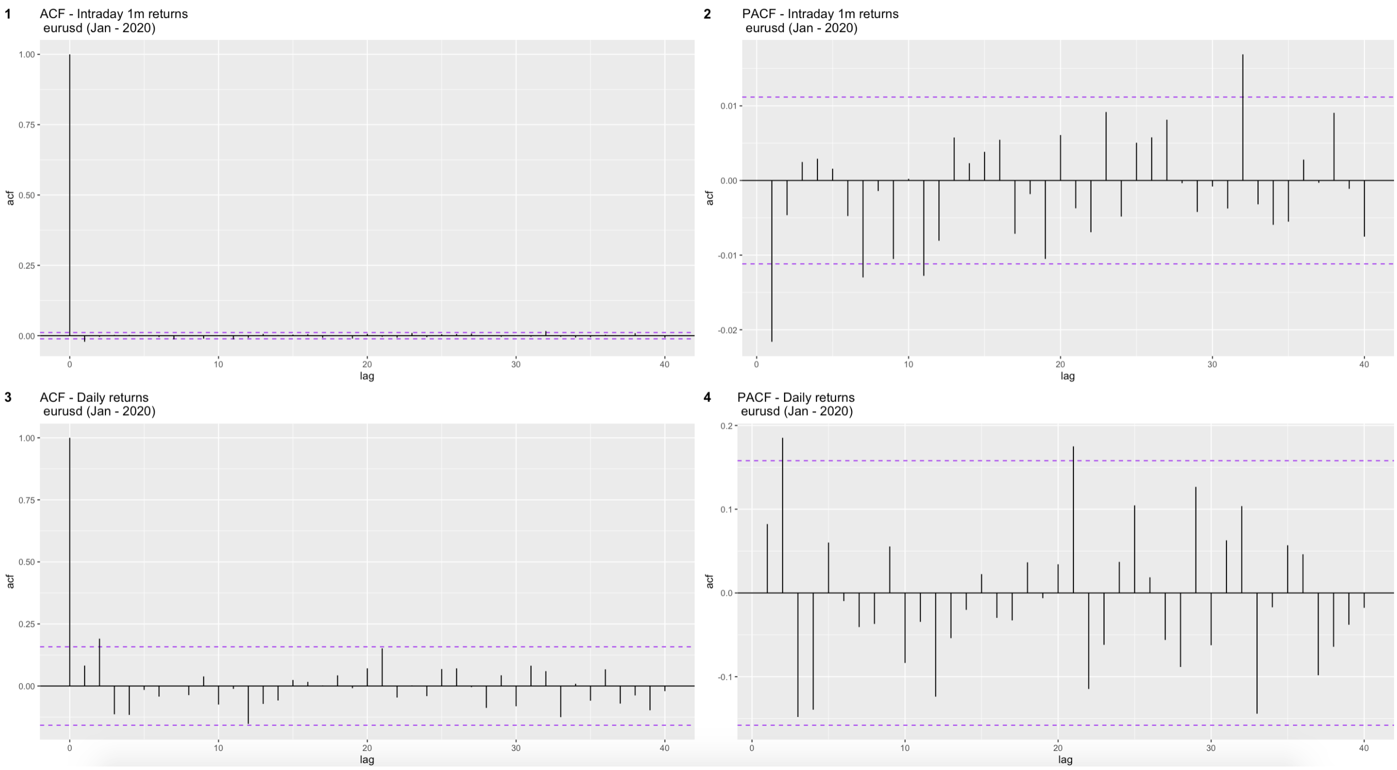


ACF



5c. Identifying the Conditional Mean Equation

- Starting with implementation of GARCH model on **conditional variance.**
- Literature suggests an Auto Regressive Integrated Moving Average (ARIMA) model, for **conditional mean.**
- Since return is stationary, the parameter *d* of ARIMA (*p, d, q*) is 0**.**
- To visually determine *p* and *q*, we should conduct a graphical analysis of **Auto Correlation Function (ACF)** and **Partial Auto Correlation Function (PACF)**



Visually Explaination

- To visually intepret both ACF and PACF and determine the order of paramters *p* an *q* as a checking of lagged dependecies:

- Suggested parameters are both 0.

Further Verification

- To further confirm the order of the mean equation, **several ARMA(p,q) models** are estimated, and the best model was chosen based:
  - **Lower** value of Akaike information criterion (AIC)
  - **Higher** value of the maximum log-likelihood
- As stated by **Mondal et al. (2014), the Box-Jenkins methodology** states that:
  - The value of p and q for an ARIMA (*p, d, q*) model should be equal to or less than 2, **or**
  - The total number of parameters should be no more than 3.
- By only checking those cases, still **ARMA(0,0) model** provided the **lowest AIC value** and the **maximum log-likelihood value** for both the 1-min return and for the daily returns series and therefore **outperforms** the other ARMA specifications for the conditional mean.

5d. Model Checking for the Mean Equation

- According to **Tsay (2005)**, there is a need to eliminate any **significant correlations in the return series** prior to fitting any GARCH-type model.
- The residuals of the mean equation are therefore tested for the presence of autocorrelations using the Ljung-Box Q test.
- All the p values were greater than 5% at 10 and 20 degrees of freedom, implying that the residuals of the mean equation are not serially autocorrelated for the two return datasets.
- At this stage, since the two return datasets exhibit stylized features such as excess kurtosis and clustering of volatility and given the adequacy of the ARMA specifications for the mean equations, the specification of GARCH models to the returns datasets is analysed.

5e Estimation of Daily Variance (h_t) Forecast

- The implementation of **MC-GARCH model** requires a **daily variance component**.
- We fit **GARCH (1, 1)** and **EGARCH (1, 1) models**, implemented under five error distributions:
  - Normal distribution
  - Student's-t distribution
  - Skewed Student's-t distribution
  - Johnson's SU distribution (JSU)
  - Generalized Error Distribution (GED)
- In the **EGARCH Model**, γ_1 as a coefficient for **asymmetric volatility**, is significant across all errors validation, indicating asymmetric GARCH might be preferred.
- Having γ_1 being positive implies that shocks including both good news and bad news which may impact the daily EUR/USD returns will affect volatility for a long period of time in the future.
- We choose the best model by comparing:
  - **Lower** value of Akaike information criterion (AIC)
  - **Higher** value of the maximum log-likelihood
- The results show that **Skewed Student's-t under GARCH(1, 1)** is preferred as the **daily variance forecast**.
- We can consider having **GED under EGARCH(1, 1)** to capture the **leverage effect** of daily return series.

	GARCH (1, 1)					EGARCH (1, 1)				
Criteria	Normal	Student's-t	Skewed Student's-t	JSU	GED	Normal	Student's-t	Skewed Student's-t	JSU	GED
AIC	-8.2838	-8.2899	-8.2795	-8.2791	-8.2850	-8.2655	-8.2759	-8.2655	-8.2645	-8.2678
BIC	-8.2159	-8.2051	-8.1777	-8.1773	-8.2002	-8.1807	-8.1741	-8.1467	-8.1457	-8.1660
Log - Likelihood	799.2450	800.8349	800.8357	800.7954	800.3596	798.4864	800.4852	800.4857	800.3909	799.7086

- This step is for reference, we should further consider their performance under MC-GARCH fitting.

5f. Fitting Performace

- Each variance model and error distributions were fitted to the **1 min EURUSD intraday dataset**.
- The result below shows the **MC-GARCH parameters estimation**.
- **The level of significance *p-values*** is provided within parentheses.
  - The null hypothesis states there is no statistical correlation between MC-GARCH model and the variable.
  - The smaller the value, the stronger evidence we should reject the null hypothesis.
  - The comments following were based on widely used 5% significance.

	MCGARCH (1, 1)									
	GARCH (1, 1)					EGARCH (1, 1)				
	Normal	Student's-t	Skewed Student's-t	JSU	GED	Normal	Student's-t	Skewed Student's-t	JSU	GED
AIC	-16.8313	-16.8622	-16.8628	-16.8651	-16.8898	-16.8243	-16.8628	-16.856	-16.8586	-16.8897
BIC	-16.8302	-16.8608	-16.8612	-16.8635	-16.8885	-16.8232	-16.8615	-16.8543	-16.857	-16.8883
Log - Likelihood	258995.4864	259471.472	259482.0167	259517.6248	259897.5035	258887.5059	259481.6814	259377.1127	259418.3038	259894.764
mu	0 (0.91795)	0 (0.420166)	0 (0.997803)	0 (0.981077)	0 (0.996791)	0 (0.96269)	0 (0.401951)	0 (0.886677)	0 (0.870832)	0 (0.996059)
omega	0.023894 (0)	0.014112 (3e-05)	0.015534 (9e-06)	0.016051 (0)	0.017429 (0)	0.015031 (0)	0.013777 (4e-06)	0.004244 (4e-06)	0.004105 (4e-06)	0.015994 (0)
alpha1	0.047257 (0)	0.057052 (0)	0.057703 (0)	0.058281 (0)	0.058762 (0)	0.043844 (0)	0.057967 (0)	0.052942 (0)	0.052849 (0)	0.060014 (0)
beta1	0.928721 (0)	0.93246 (0)	0.930298 (0)	0.92915 (0)	0.928777 (0)	0.941167 (0)	0.932029 (0)	0.946058 (0)	0.94615 (0)	0.929006 (0)
shape	NA (NA)	5.976128 (0)	0.988715 (0)	-0.034417 (0.077307)	1.088898 (0)	NA (NA)	5.854991 (0)	0.991297 (0)	-0.025152 (0.167561)	1.085905 (0)

i. Overall Observation

- All the parameter estimates are **statistically significant at 5% level**, except for the conditional mean (mu), which is insignificant at 5% level across all innovations for the MC-GARCH model.
- The **skewness** parameter is insignificant for the **JSU innovation**.
- Almost all the parameter estimates being statistically significant gives an indication that the MC-GARCH models are correctly specified.

ii. Observation by Parameters

ARCH and GARCH Parameters (α_1 and β_1)

- The statistical significance of the **ARCH parameter (α_1)** and **GARCH parameter (β_1)** for all innovations of the MC-GARCH model suggests that:
  - The **lagged conditional variance** and **lagged squared disturbance** have an impact on the current conditional variance.
- This simply implies that **news about volatility** from the previous periods have an **explanatory power** on the current volatility.
- Moreover, the high significance of the parameter α_1 validates the presence of volatility clustering in the dataset.

Shape (ν)

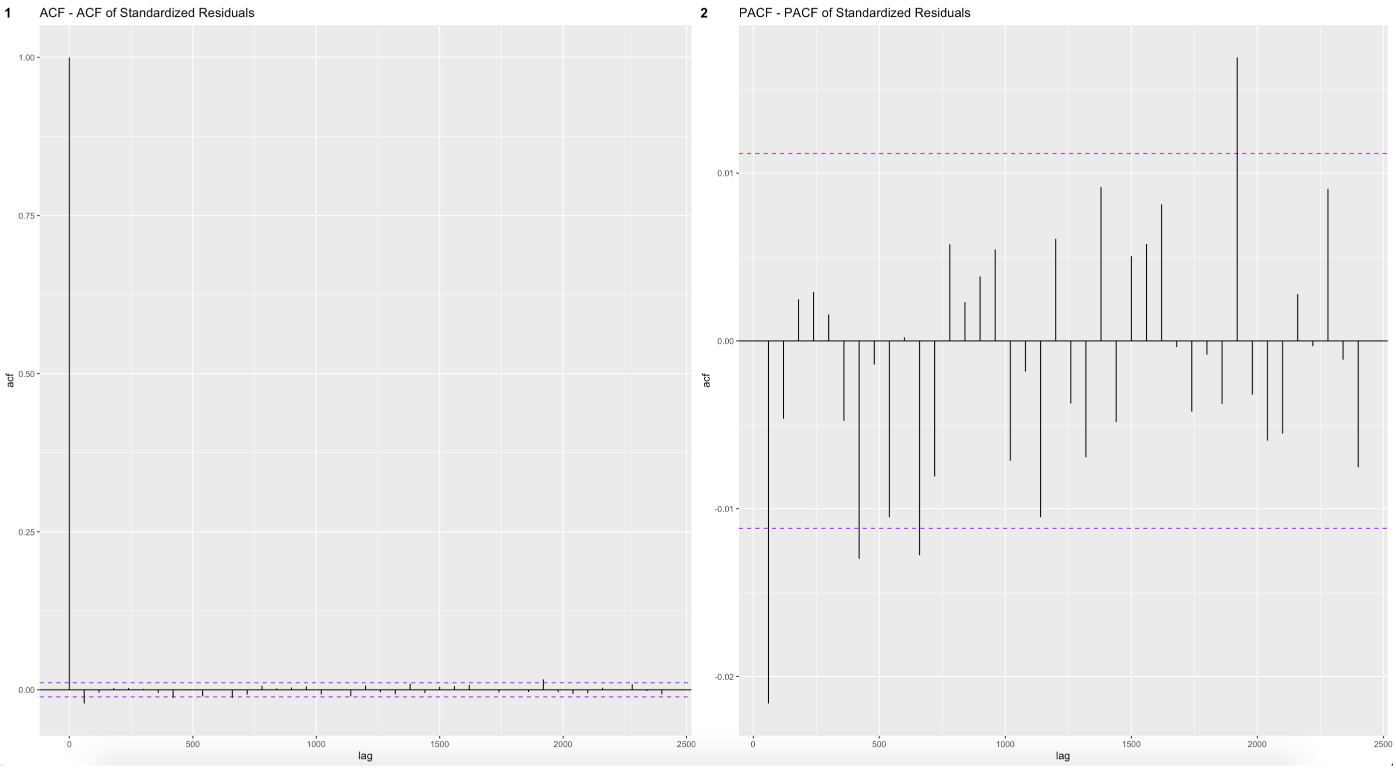
- The **shape** ν, parameter being **highly statistically significant**:
  - Greater than 4 for the Student's-t
  - Less than 2 for the GED innovation

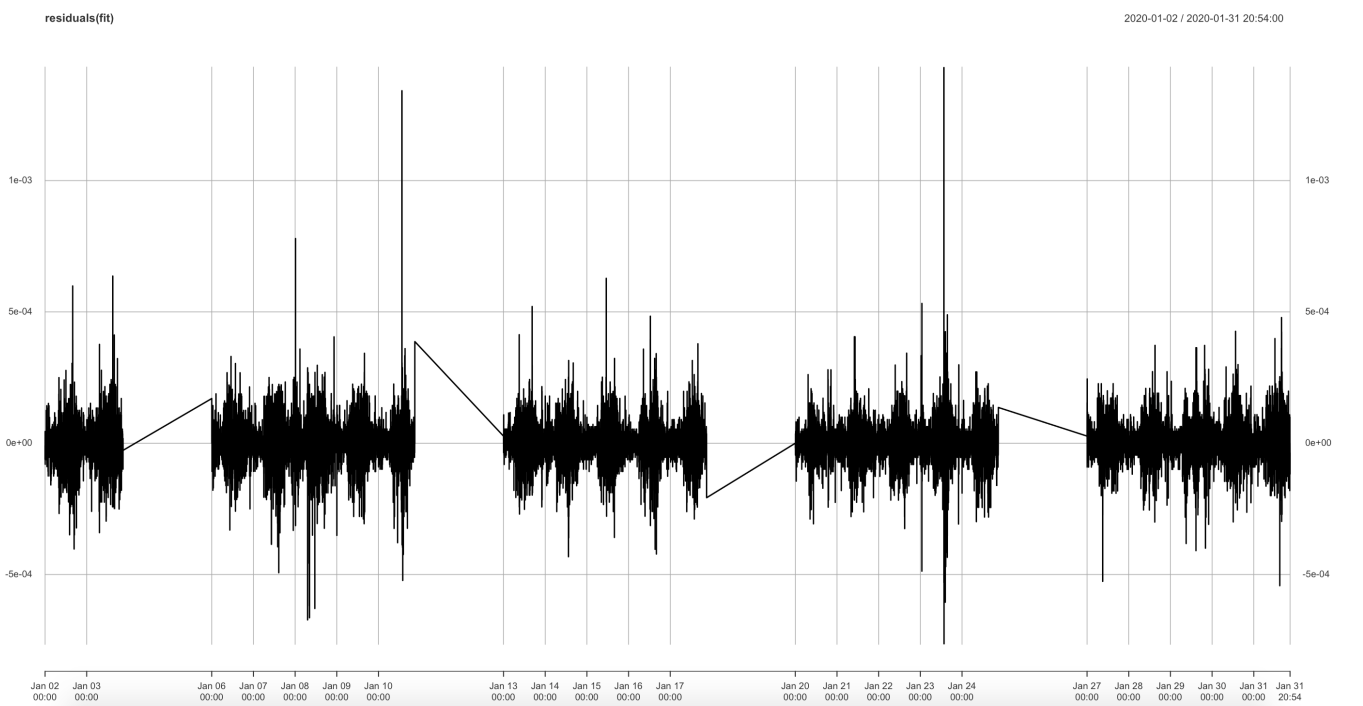
- These confirms the **presence of thick tails (leptokurtic)**, as shown by the excess kurtosis in the return dataset of the 1-min EUR/USD returns.
- These results suggest that a **non-normal innovation** to be a more suitable candidate for the MC-GARCH model.

iii. Chosen Model

- We choose MC-GARCH model with **EGARCH(1, 1)** as the daily variance model to **capture the leverage effect feature of daily return series** and **GED** as the error distribution.

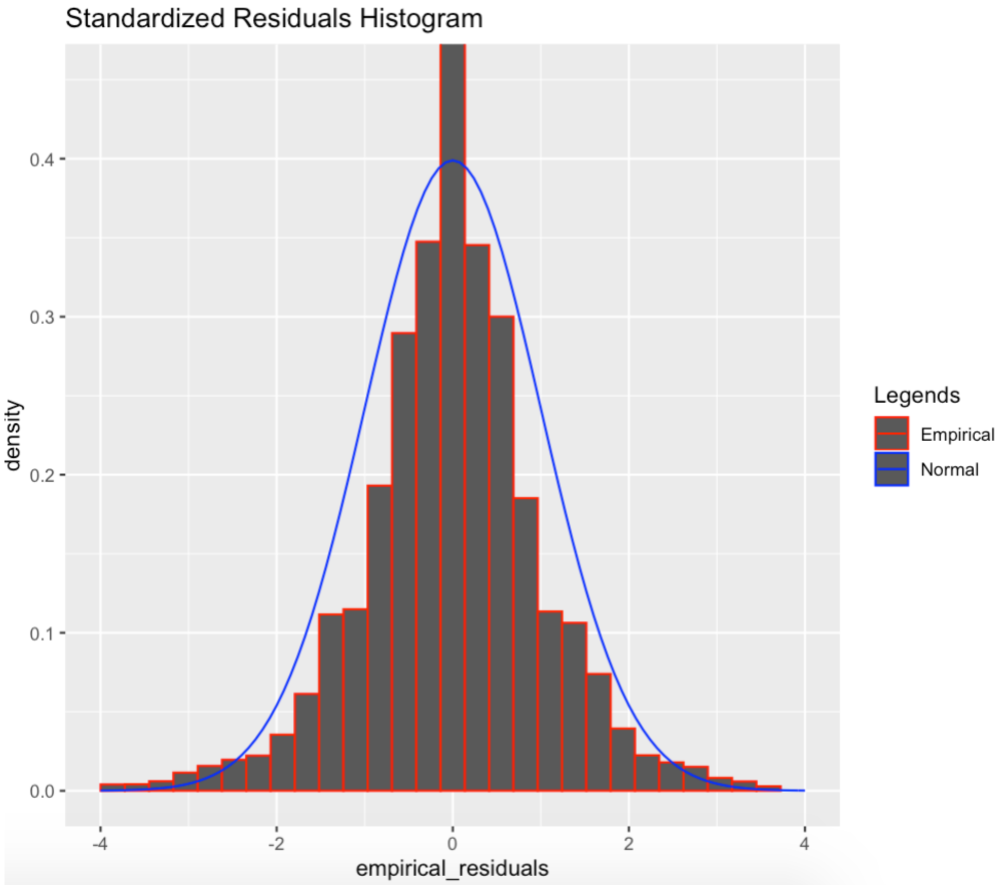
iv. Residuals





- ACF and PACF of the residuals show **no serial correlation between lags**, which behave like a **noise process**.

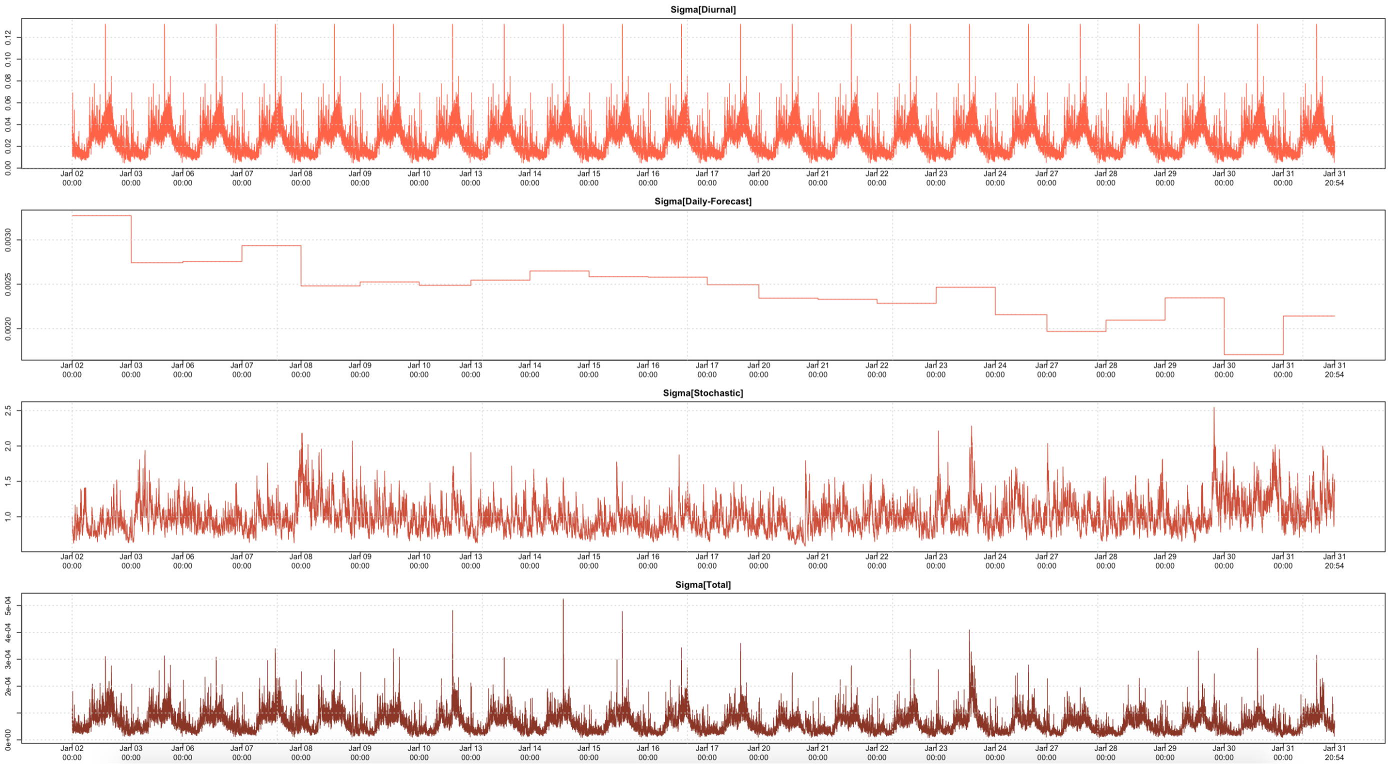
v. Histogram of residuals (Normality)



- Again the residuals is visually following a normal distribution.

vi. Variance Decompostion

- The following chart is for:
  - Diurnal Component d_t to cater intraday seasonality,
  - Daily volatility Component h_t,
  - Intraday Volatility Component q_t,
  - Total Composite Volatility σ_t



5g. Intraday VaR Forecast

- The 99% intraday VaR is forecasted using the MC-GARCH models on the 1-min intraday return series.
- A rolling backtest procedure is then undertaken on the out-sample period and a moving window of 1 day will be used in the VaR backtesting procedure.
- The backtesting period the last 2000 1 min intraday return in January 2020.
- We should introduce some VaR statistical testing such as:
  - Kupiecs's Unconditional Coverage Test
  - A Duration-Based Approach to VaR Backtesting
  - Asymmetric VaR Loss Function

Kupiec's Unconditional Coverage Test

- The Kupiec's test was developed by **Kupiec (1995)** and is the most famous VaR test that is based on failure rates. * It is also known as the **proportion of failures (POF) test.**
- The null hypothesis of the test assumes that the number of exceptions follows a **binomial distribution.**
- The null hypothesis for the test is as follows:

$$H_0 = p = \hat{p} = \frac{x}{T}$$

- where:
  - T is the number of observations
  - x is the number of exceptions.
- The test is in fact a likelihood ratio test where the test statistics are as follows:

$$LRPOF = -2ln(\frac{(1-p)^{T-x}p^x}{[1-(x/T)]^{T-x}(x/T)^x})$$

- Under the null hypothesis, the LRPOF is asymptotically chi-square distributed with one degree of freedom.

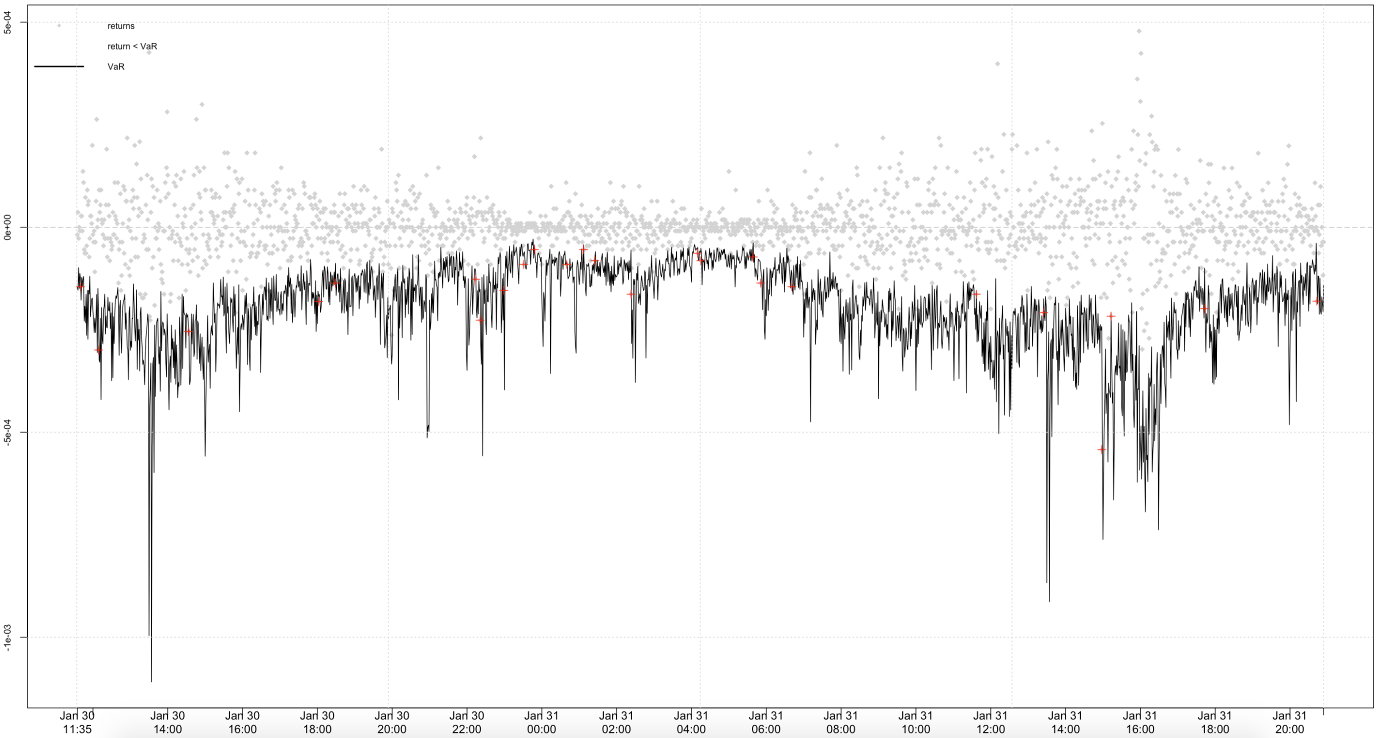
Results:

	GARCH (1, 1)				
Criteria	**Normal**	**Student's-t**	**Skewed Student's-t**	**JSU**	**GED**
Expected VaR Exceedances	20.0	20.0	20.0	20.0	20.0
Actual VaR Exceedances	23.0	23.0	24.0	24.0	24.0
Actual (%)	1.10%	1.10%	1.20%	1.20%	1.20%
p - values	0.5100	0.5100	0.3830	0.3830	0.3830

EGARCH (1, 1)				
Normal	**Student's-t**	**Skewed Student's-t**	**JSU**	**GED**
20.0	20.0	20.0	20.0	20.0
26.0	25.0	21.0	21.0	25.0
1.30%	1.20%	1.10%	1.10%	1.20%
0.1970	0.2790	0.8240	0.8240	0.2790

- We have all of the models passes (being greater than 5% significance level).
- This indicates the null hypothesis cannot be rejected.

Back-tested VaR



- The spikes in VaR forecast is due to seasonal component during opening of each trading day, and ending of London trading session.