Forecasting realized volatility using GARCH-type models: estimation and prediction

July 2022

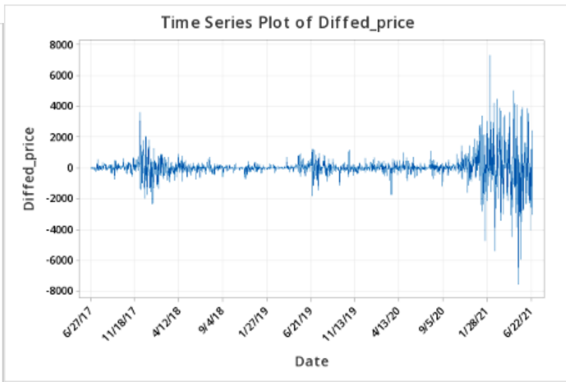
1 Context and Data Source

Bitcoin is a decentralized digital currency that can be transferred on the peer-to-peer bitcoin network. Bitcoin transactions are verified by network nodes through cryptography and recorded in a public distributed ledger called a blockchain. Given the present and past price, how much it will be is a question of interest for investors. The goal of this project is to use ARIMA-GARCH-type methods to forecast the time series data "Adjusted Close Price for Bitcoin to USD" (logged) based on four-year daily data from June 27th, 2017 to June 27th, 2021 (deleting several most extreme observations from March 12th to March 18th 2020) with 1456 observations in total provided by Yahoo: `https://finance.yahoo.com/quote/BTC-USD/history?period1=1561593600&period2=1624752000&interval=1d&filter=history&frequency=1d&includeAdjustedClose=true`

2 Data Description, ARIMA Identification, and Parameter Estimation

First, I looked at the time series plots for Bitcoin-USD and the first difference of Bitcoin-USD.

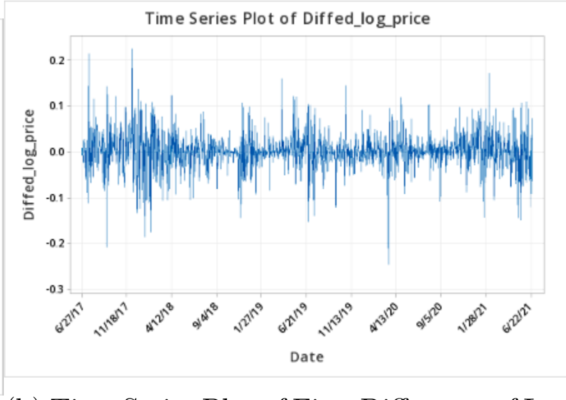
(a) Time Series Plot of BTC-USD



(b) Time Series Plot of First Difference of BTC-USD

From both plots, we notice that there is level-dependent volatility, and there is also an exponential trend of the price of BTC-USD over time, which means I should probably take logs. Here are the time series plots for log BTC-USD and the first difference of log BTC-USD.

(a) Time Series Plot of Log BTC-USD



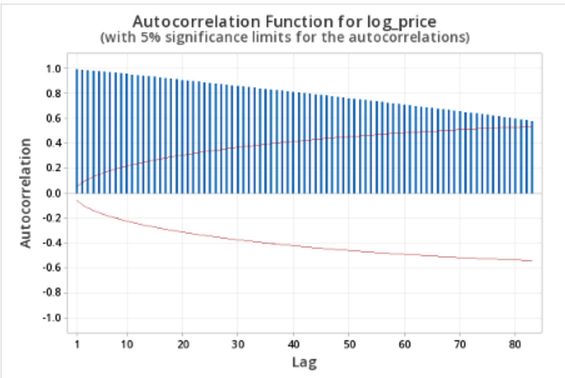
(b) Time Series Plot of First Difference of Log BTC-USD

After taking logs, I removed the level-dependent volatility as well as linearize the trend. I will work with the logged data rather than the original price of BTC-USD.

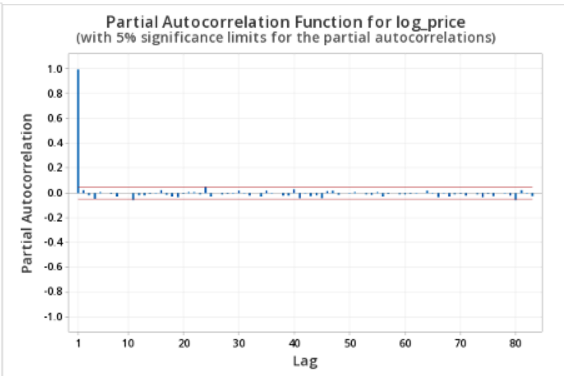
Overall, log BTC-USD exhibits an increasing trend over time. From the first difference of the log BTC-USD, it is also worth noticing that there is volatility clustering. In particular, I witness the stock was very volatile around March 20th, 2020, which comes hand-in-hand with the coronavirus outbreak, affecting global markets and driving investors towards the safety of cash.

3 ARIMA Model Identification

To start with, I have the following ACF and PACF plots for log BTC-USD.

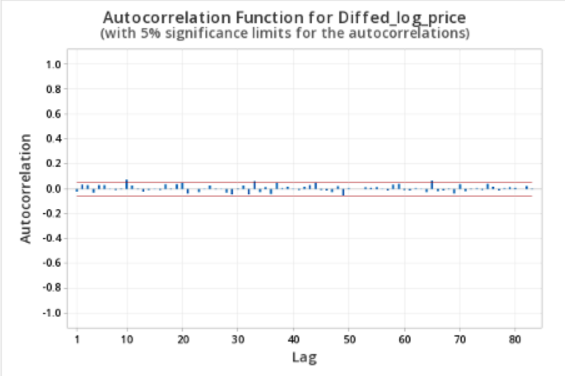


(a) ACF Plot of Log BTC-USD

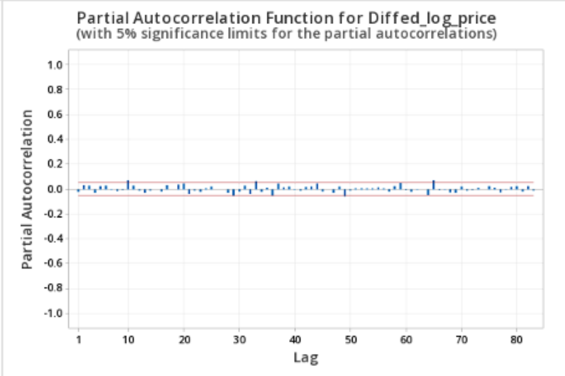


(b) PACF Plot of Log BTC-USD

The series do not appear to be stationary. The autocorrelation of log BTC-USD keeps hanging for many lags, and PACF cuts off beyond lag 1 whose partial autocorrelation is around 1. These are evidence that the series are not stationary, which indicates that we should try differencing the data. After differencing the log BTC-USD time series, I got the following ACF and PACF plots.



(a) ACF Plot of First Difference of Log BTC-USD



(b) PACF Plot of First Difference of Log BTC-USD

In both the ACF and PACF plots, the autocorrelation or partial autocorrelation are not statistically significant in many lags. To make sure the time series of first difference of log BTC-USD is stationary, I difference this series again and get the following ACF plot.

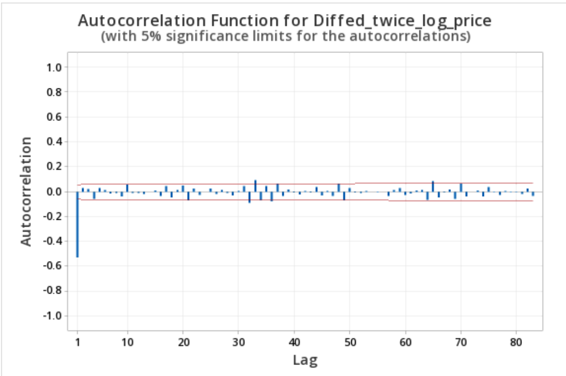


Figure 5: ACF Plot of the Second Difference of Log BTC-USD

From the ACF plot of the second difference of log BTC-ISD, the lag 1 autocorrelation is statistically negative, which is a sign of overdifferencing. Therefore, we should only take the difference once and the d in my ARIMA model should be 1.

The ACF for the first difference of the logged data suggests q might be 0. The PACF for the first difference of the logged data suggests p=0. However, to determine which model to use ultimately, I will loop over 0 to 2 for both p and q and to identify their values based on AICc. Based on the time series plot of differenced log BTC-USD, I choose not to include a constant in the ARIMA model.

For ARIMA(p,1,q) models without constant, the following formula should be used to calculate AICc:

$$AICc = Nlog\frac{SS}{N} + \frac{2(p+q+1)}{N-p-q-2} \cdot N \tag{1}$$

where N is 1456-1=1455, p and q are subject to change

p	q	SS	AICc
0	0	2.55364	-9230.323466
0	1	2.55747	-9226.137351
0	2	2.55359	-9226.338167
1	0	2.55741	-9226.171487
1	1	2.55675	-9224.538756
1	2	Error	None
2	0	2.55401	-9226.098877
2	1	2.54798	-9227.527138
2	2	2.54750	-9225.787441

Table 1: ARIMA$(p, 1, q)$ Models with p, q less than 2

The smallest AICc is -9230.323466 given by the model ARIMA (0,1,0) without constant. Let the log BTC-USD be $\{x_t\}$, and the first difference of log BTC-USD be

$$y_t = x_t - x_{t-1} \tag{2}$$

Given the parameter estimation, we have

$$y_t = \epsilon_t$$

which implies

$$x_t - x_{t-1} = \epsilon_t \tag{3}$$

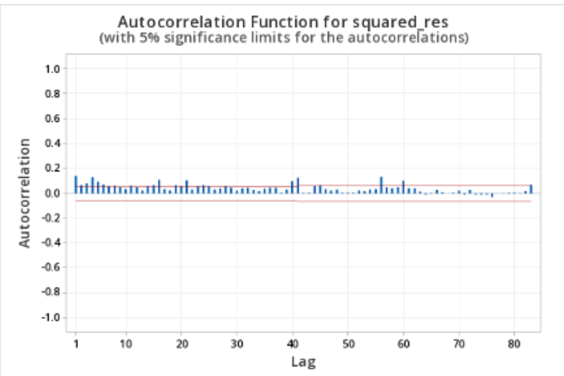
therefore, the complete form of the fitted model is

$$x_t = x_{t-1} + \epsilon_t \tag{4}$$

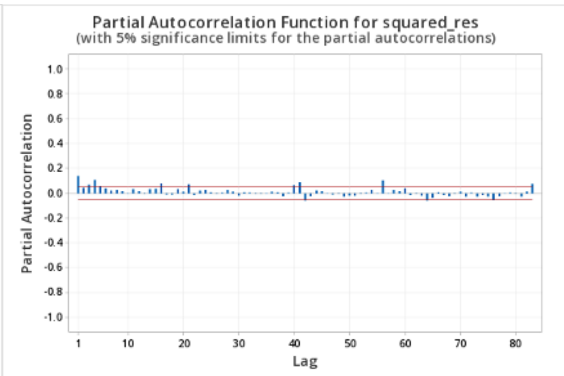
Therefore, we know log BTC-USD series is a random-walk and the best forecast for tomorrow is the price of today. Based on our data and the above analysis, we can conclude that the bitcoin market is an efficient market.

4 Residuals and Heteroskedasticity

Given ARIMA(0,1,0), the residuals are actually the first difference of log BTC-USD. The plot of residuals and its ACF and PACF plots are shown above. ACF and PACF of both the squared residuals are as follows.



(a) ACF of Squared Residuals



(b) PACF Plot of Squared Residuals

In general the residuals are approximately uncorrelated. However, the residuals are not independent. First, we see volatility clustering. Second, if the residuals are independent, the squared residuals would also be independent so that there won't be autocorrelation or partial autocorrelation between squared residuals, which contradicts with the ACF and PACF plots of our squared residuals. Therefore, the residuals are not independent. Instead, they show evidence of heteroskedasticity.

5 GARCH Models

Next, we use the most fundamental GARCH to model the conditional distribution of residuals from the ARIMA model. The followings are the log likelihood and corresponding AICc for GARCH(p,q) models.

AICc is calculated using the following formula:

$$AICc = -2 \cdot loglikelihood + \frac{2(q+1)}{N-q-2} \cdot N \tag{5}$$

p	q	log likelihood	AICc
1	1	2641.998	-5277.97946
2	1	2640.041	-5272.05441
1	2	2627.545	-5247.06241
2	2	2631.953	-5253.86459
2	3	2624.325	-5236.59199

Table 2: GARCH(p, q) Models with Different p, q

$v(m/s)$	$Drag(N)$	$Radius(m)$
0.1	2.6904	5.3428
0.2	5.3948	5.3498
0.3	8.1134	5.3568
0.4	10.8462	5.3638
0.5	13.5936	5.3709
0.6	16.3549	5.3779
0.7	19.1311	5.3850
0.8	21.9218	5.3921
0.9	24.7271	5.3992
1.0	27.5478	5.4064

The AICc for GARCH (1,1) model is -5277.97946, which is the smallest AICc in models above. Therefore, we would select the GARCH (1,1) model as our preferred one. The summary of the model is as follows:

```
Model:
GARCH(1,1)

Residuals:
     Min       1Q   Median       3Q      Max 
-6.68061 -0.41629  0.04343  0.48869  6.16733 

Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)    
a0 6.762e-05   8.600e-06    7.863 3.77e-15 ***
a1 7.814e-02   8.469e-03    9.227  < 2e-16 ***
b1 8.852e-01   1.207e-02   73.315  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Diagnostic Tests:
        Jarque Bera Test

data:  Residuals
X-squared = 1512.5, df = 2, p-value < 2.2e-16
```

The parameters in the model are all highly statistically significant since their p-value (and half of p-value) are all much smaller than 0.05. Therefore, the complete form for the model should be

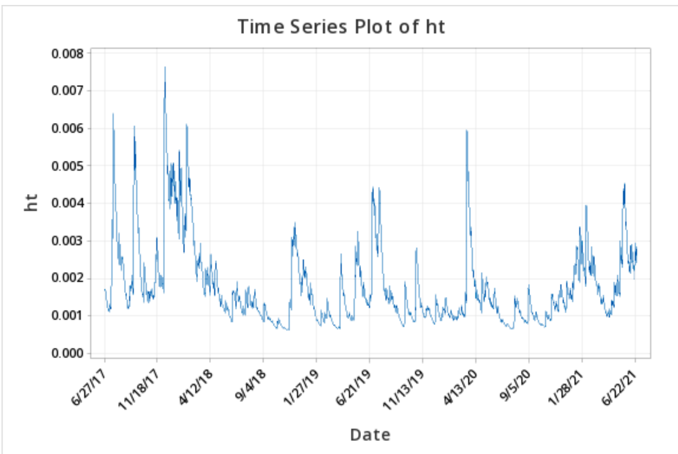
$$h_t = 6.762 \cdot 10^{-5} + 0.07814\epsilon_{t-1}^2 + 0.8852h_{t-1}$$

where $\{\epsilon_t\}$ is the residuals from the fitted ARIMA model.

The unconditional (marginal) variance of the shocks

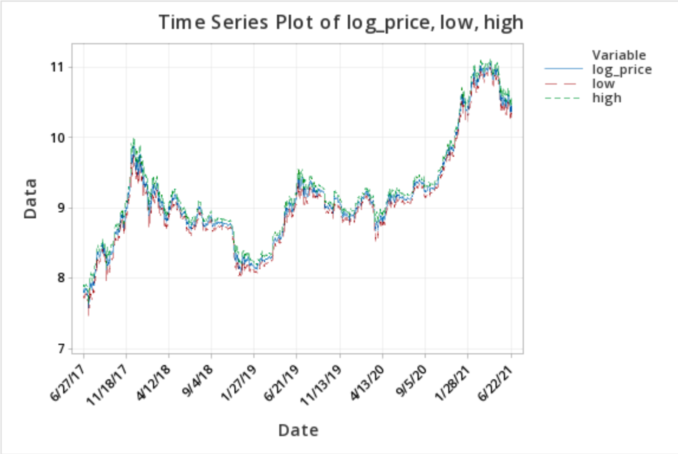
$$Var(\epsilon_t) = \frac{a_0}{1 - a_1 - b_1} = \frac{6.762 \cdot 10^{-5}}{1 - 0.07814 - 0.8852} = 1.8445 \cdot 10^{-3}$$

6 Conditional Variance Plot



The burst of high volatility is seen around March 20 2020. This highly volatile period agrees with the one found from examination of the time series plot of the log BTC-USD itself, which reflects the coronavirus outbreak.

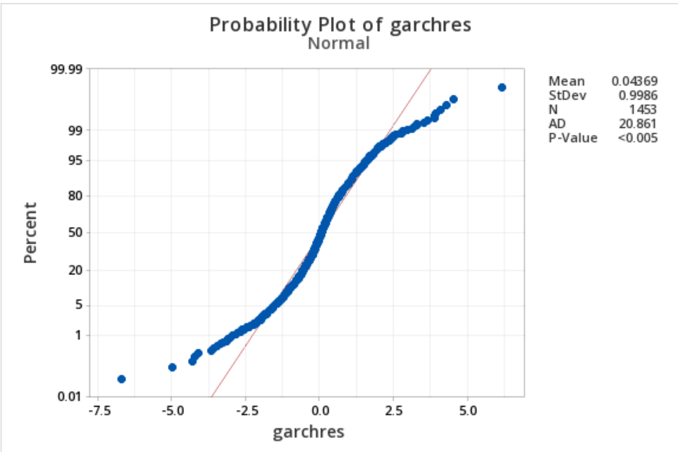
7 Forecast Interval Plot



It seems that most of the true values of log BTC-USD are contained in the 95% forecast interval. Therefore, the forecast interval is accurate. This forecast is useful in the sense that the interval is not large and it can make relatively good predictions about log BTC-USD. However, it might not be that useful practically because the performance here might be better than that in the actual forecasting context since we get the estimates based on the entire dataset rather than up to the time the forecast interval is constructed.

8 GARCH Residuals

The following plot shows the normality test for the GARCH residuals.



The p-value for the normality test is less than 0.005 so that we should reject the null hypothesis that the residuals are normally distributed. From the normality plot, the model does not seem to have adequately described the leptokurtosis in the data. We still need to use advanced GARCH models or other distribution assumptions to fit this data.

9 Frequency of Predicting Failure

There are 79 times out of 1454 data values when the 95% forecast interval does not cover the actual value, which accounts for $79/1454 = 5.433\%$ in total. Therefore, the in-sample accuracy of the prediction is 94.567%.

10 Common Distribution Methods and Characteristics

The choice of the conditional distribution of residuals in the model plays a crucial role in the fitting effect and explanatory power of the model. Financial time series mostly show characteristics of high kurtosis and heavy tails, and have volatility clustering and long memory, etc. In this research, we assumed that the ARIMA residuals conform to the normal distribution, the Student-t distribution, and the generalized error distribution (GED), and employed the maximum likelihood method in R for parameter estimation respectively. Bollerslev and Wooldridge (1992) states that the proposed maximum likelihood estimation satisfies the collinearity and asymptotic normality. The log-likelihood function (see section 10.1 and 10.2) of the proposed maximum likelihood estimate of the model differs depending on the choice of the conditional distribution, which will be analyzed in the following subsections.

10.1 Log-likelihood Function

Suppose a random variable Ξ and we observe a realization (a vector) of the random variable $\xi = [x_1 ... x_n]$, whose entries $x_1, ..., x_n$ are draws from certain unknown distribution. The distribution of Ξ belongs to a parametric family: there is a set $\Theta \subseteq \mathbb{R}^p$ of real vectors (called the parameter space) whose elements (called parameters, denoted by θ. For example, for normal distribution, the parametric vector is $\theta = [\mu \ \ \sigma^2]$) are put into correspondence with the distributions that could have generated Ξ.

- If Ξ is a continuous random vector, its joint probability density function belongs to a set of joint probability density functions $f(\xi; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$
- If Ξ is a discrete random vector, its joint probability mass function belongs to a set of joint probability mass functions $p(\xi; \theta) = \prod_{i=1}^{n} p(x_i; \theta)$

Denote the likelihood function $L(\theta; \xi)$ by the joint probability mass (or density) function as a function of θ for fixed ξ, namely:

- $L(\theta; \xi) = f(\xi; \theta)$ for Ξ continuous
- $L(\theta; \xi) = p(\xi; \theta)$ for Ξ discrete

Given all these elements, the log-likelihood function is the function $l(\theta; \xi)$ defined by $l(\theta; \xi) = \ln[L(\theta; \xi)]$.

10.2 How the Log-likelihood is Used

The log-likelihood function is typically used to derive the maximum likelihood estimator of the parameter θ. The estimator $\hat{\theta}$ is obtained by solving:

$$\hat{\theta} = \arg \underset{\theta \in \Theta}{max}\, l(\theta; \xi) \tag{6}$$

That is, by finding the parameter $\hat{\theta}$ that maximizes the log-likelihood of the observed sample ξ. This is the same as maximizing the likelihood function $L(\theta; \xi)$ because the natural logarithm is a strictly increasing function.

However, since the computer optimization algorithms are often written as minimization algorithms, equation (6) is equivalently written in terms of the negative log-likelihood as $\hat{\theta} = \arg \underset{\theta \in \Theta}{min}[-l(\theta; \xi)]$ before being solved numerically on computers.

10.3 Normal Distribution

The normal distribution is a very versatile distribution with many advantages. Traditional economists have proposed the assumption that the residual approximately obeys the normal distribution, which is widely accepted by the industry. If the probability density function of a random variable X is:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{1}{2\sigma^2}(x-a)^2), -\infty < x < +\infty \tag{7}$$

Then the log-likelihood functions is:

$$L_N = -\frac{1}{2}T\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log{\sigma_t}^2 - \frac{1}{2}\sum_{t=1}^{T}\frac{{\epsilon_t}^2}{{\sigma_t}^2} \tag{8}$$

However, the normal distribution has some limitations in the real stock market. For example, the sequence of financial asset returns exhibits high-kurtosis, thick-tail and leverage effects. And because the speed of information dissemination, the speed of investors' reaction, investors' preference for risk are all different, the normal distribution assumption is not as credible as one might think.

10.4 Student-t Distribution

Let the random variable X obey the standard normal distribution $N(0,1)$, Y obey the chi-square distribution $\chi^2(n)$, X and Y are independent, then the distribution of $Z = \frac{X}{\sqrt{Y/n}}$ is said to be a t-distribution with degree of freedom n, denoted as $Z \sim t(n)$. If the residual series obey the conditional distribution as a t-distribution with degree of freedom n, the probability density function is:

$$f_Z(x,n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}(1+\frac{x^2}{n})^{-\frac{n+1}{2}}, -\infty < x < +\infty \tag{9}$$

Then the log-likelihood function is:

$$L_t = T\log[\frac{\Gamma[\frac{n+1}{2}]}{\sqrt{\pi}\sqrt{n-2}\cdot\Gamma(\frac{n}{2})}] - \frac{1}{2}\sum_{t=1}^{T}\log({\sigma_t}^2) - (\frac{n+1}{2})\cdot\sum_{t=1}^{T}\log[1+\frac{\epsilon_t^2}{\sigma_t^2(n-2)}] \tag{10}$$

The density function of the t-distribution is similar to the standard normal distribution in shape and is symmetric about the vertical axis, but the kurtosis is lower and the tail is heavier than the standard normal distribution. The smaller the degree of freedom n, the lower the kurtosis and the thicker the tail of the curve will be, showing a flatter pattern; the larger the degree of freedom n, the higher the kurtosis, and the curve will get closer to the standard normal distribution. Therefore, for a small degree of freedom, it can be used as a model for financial returns exhibiting excessive kurtosis, enabling a more realistic calculation of the Value at Risk (VaR).

10.5 Generalized Error Distribution (GED)

The Generalized Error Distribution (GED) is a symmetric single-peaked function in the exponential family of functions and is a flexible probability distribution function whose kurtosis tends to infinity with respect to the parameters. The generalized error distribution has a heavier tail than the standard normal distribution and is therefore widely used to describe price fluctuations in financial markets. Subbotin

(1923) proposed that GED includes the special cases of normal distribution, the Laplace distribution and the uniform distribution. The parameter can be adjusted to fit different forms. If the random variable X obeys a GED with parameter v, its density function can be expressed as:

$$f(x,v) = \frac{v}{\lambda_v \cdot 2^{(\frac{v+1}{v})}\Gamma(\frac{1}{v})} \cdot \exp(-\frac{1}{2}|\frac{x}{\lambda_v}|^v) \tag{11}$$

where λ_v is denoted as:

$$\lambda_v = [\frac{2^{(-\frac{2}{v})} \cdot \Gamma(\frac{1}{v})}{\Gamma(\frac{3}{v})}]^{\frac{1}{2}} \tag{12}$$

Then the log-likelihood function is:

$$L_{GED} = \sum_{t=1}^{T}[\log(\frac{v}{\lambda_v}) - \frac{1}{2}|\frac{x_t}{\lambda_v}|^v - (1+\frac{1}{v})\log 2 - \log\Gamma(\frac{1}{v} - \frac{1}{2}\log(\sigma_t^2))] \tag{13}$$

GED is useful when the errors around the mean or in the tails are of special interest. It is very flexible and we can fit thick-tailed phenomenon of different degrees by adjusting the size of the parameter v in equation (11). $v = 2$ shows a normal distribution; when $v < 2$ the tail will be heavier and thinner vice versa.

11 GARCH Variations and Comparison

Model	Distribution Method	Log-likelihood	AICc
GARCH(1, 1)	normal	2644.23	-5282.44
GARCH(2, 1)		2640.04	-5272.05
GARCH(1, 2)		2627.55	-5247.06
GARCH(2, 2)		2631.95	-5253.86
GARCH(2, 3)		2624.32	-5236.59
GARCH(1, 1)	student-t	2807.04	-5608.06
	GED	2806.53	-5607.05
EGARCH(1, 1)	normal	2645.64	-5285.27
	student-t	2817.66	-5629.30
	GED	2809.48	-5612.93
TGARCH(1, 1)	normal	2649.40	-5292.79
	student-t	2816.73	-5627.44
	GED	-3180.31	6366.64
IGARCH(1, 1)	normal	2632.22	-5258.43
	student-t	2807.40	-5608.79
	GED	2806.51	-5607.01

Table 3: GARCH Variations with Different Distribution Method

The lowest AICc is given by the model EGARCH(1,1) with student-t distribution.

12 Optimal Model Parameter Estimation

The following pictures illustrate parameter estimation of the model EGARCH(1,1) with student-t distribution given by R.

```
*---------------------------------*
*          GARCH Model Fit        *
*---------------------------------*

Conditional Variance Dynamics
-----------------------------------
GARCH Model     : eGARCH(1,1)
Mean Model      : ARFIMA(0,0,0)
Distribution    : std

Optimal Parameters
------------------------------------
        Estimate  Std. Error   t value Pr(>|t|)
omega  -0.012469    0.010204  -1.22207  0.22168
alpha1  0.005296    0.016930   0.31282  0.75442
beta1   0.997737    0.002047 487.46134  0.00000
gamma1  0.226841    0.020519  11.05520  0.00000
shape   2.776696    0.221638  12.52805  0.00000

Robust Standard Errors:
        Estimate  Std. Error   t value Pr(>|t|)
omega  -0.012469    0.014214  -0.87725  0.38035
alpha1  0.005296    0.017872   0.29633  0.76698
beta1   0.997737    0.002050 486.63860  0.00000
gamma1  0.226841    0.018876  12.01715  0.00000
shape   2.776696    0.213363  13.01397  0.00000

LogLikelihood : 2817.66

Information Criteria
------------------------------------

Akaike       -3.8689
Bayes        -3.8507
Shibata      -3.8689
Hannan-Quinn -3.8621
```

```
Weighted Ljung-Box Test on Standardized Residuals
------------------------------------
                        statistic p-value
Lag[1]                      2.335 0.12650
Lag[2*(p+q)+(p+q)-1][2]     3.540 0.10025
Lag[4*(p+q)+(p+q)-1][5]     6.849 0.05642
d.o.f=0
H0 : No serial correlation

Weighted Ljung-Box Test on Standardized Squared Residuals
------------------------------------
                        statistic p-value
Lag[1]                   0.002223  0.9624
Lag[2*(p+q)+(p+q)-1][5]  2.704574  0.4637
Lag[4*(p+q)+(p+q)-1][9]  7.260082  0.1783
d.o.f=2

Weighted ARCH LM Tests
------------------------------------
            Statistic Shape Scale P-Value
ARCH Lag[3]    0.4022 0.500 2.000 0.52598
ARCH Lag[5]    7.4346 1.440 1.667 0.02743
ARCH Lag[7]    9.0977 2.315 1.543 0.02970

Nyblom stability test
------------------------------------
Joint Statistic:  0.8588
Individual Statistics:
omega  0.1998
alpha1 0.1873
beta1  0.2060
gamma1 0.1566
shape  0.2530


Asymptotic Critical Values (10% 5% 1%)
Joint Statistic:         1.28 1.47 1.88
Individual Statistic:    0.35 0.47 0.75

Sign Bias Test
------------------------------------
                   t-value   prob sig
Sign Bias           0.1068 0.9150
Negative Sign Bias  0.5907 0.5548
Positive Sign Bias  1.0869 0.2772
Joint Effect        1.5536 0.6700


Adjusted Pearson Goodness-of-Fit Test:
------------------------------------
  group statistic p-value(g-1)
1    20     23.52      0.21505
2    30     29.91      0.41868
3    40     45.64      0.21542
4    50     62.57      0.09212


Elapsed time : 2.671799
```

13 The Principle of MCMC Algorithm

The MCMC algorithm is based on the principle that by building a Markov chain with $\pi(x)$ as a smooth distribution, then sampling $\pi(x)$, one obtains a series of samples $\{x^{(1)}, ..., x^{(n)}\}$ with respect to $\pi(x)$, and take the Monte Carlo integral over the state space for the sample. The expectation of any function $f(x)$ can be estimated using the Monte Carlo integral.

First, based on the sample obtained from the sampling method, taking mean of $f(x^{(i)})$ we have:

$$E[f(x^{(i)})] = \frac{1}{n}\sum_{i=1}^{n} f(x^{(i)})$$

The expectation of $f(x)$:

$$E[f(x)] = \int f(x)\pi(x)\mathrm{d}x$$

According to the Law of Large Numbers, since the elements of the sample $x^{(1)}$, ..., $x^{(n)}$ are independent, so when n goes to infinity, we have:

$$E[f(x^{(i)})] \xrightarrow{L.L.N} E[f(x)]$$

In other words, if the samples size is large enough, we can get the formula for estimating $E[f(x)]$ using $E[f(x^{(i)})]$:

$$E[f(x)] \approx \frac{1}{n}\sum_{i=1}^{n} f(x^{(i)})$$

$$\mathcal{L} = -\log\frac{\exp(\boldsymbol{q_j}\cdot\boldsymbol{k_j}/\tau)}{\exp(\boldsymbol{q_j}\cdot\boldsymbol{k_j}/\tau) + \sum_i \exp(\boldsymbol{q_j}\cdot\boldsymbol{k_j})}$$

$$F_{net} = \rho_{ColdAir}\cdot(\frac{4}{3}\pi R^3)g - \frac{1}{2}C\rho_{ColdAir}(\pi R^2)v^2 - (M_{human+basket} + \rho_{HeatedAir}\cdot(\frac{4}{3}\pi R^3))g$$

$$F_{net} = 4.516R^3 - 0.943v^2R^2 - 686$$