

Extreme Value Analysis of Egg Sales



December 16th 2024

Tim Bättig (374333) – Giammarco Bozzelli (376379) – Balthazar Duc (369223) –
Jakob Theo Andreas Frühauf (369224) – Alex Pinel (310388)

Group 6

Contents

1 Practical 1	1
1.1 Part 1: Financial Returns and Normality	1
1.2 Part 2: Financial Time Series, Heteroscedasticity and the Random Walk Hypothesis	2
1.3 Part 3: Dependence Between Time Series	3
2 Practical 2	4
2.1 Part 1: Block Maxima Approach	4
2.2 Part 2: Peaks-Over-Threshold (POT) Approach	5
2.3 Part 3: Clustering and Seasonal Variations	6
3 Practical 3	7
3.1 Data	7
3.2 Extreme Value Analysis	8
3.2.1 Block Maxima Approach	8
3.2.2 POT Approach	9
3.2.3 Clustering and Holidays	10

1 Practical 1

1.1 Part 1: Financial Returns and Normality

- a) To test for stationarity in the raw Bitcoin prices, we perform an Augmented Dickey-Fuller (ADF) test. The test yields a p-value of 0.3885, which is greater than 5%. This indicates that we cannot reject the null hypothesis that the time series is non-stationary. Therefore, we conclude that the raw Bitcoin price series is not stationary.
- b) On this new data, we conduct another ADF test and obtain a p-value of 0.01. With a 5% threshold, we can reject the null hypothesis that the time series is non-stationary. Thus, we conclude that the negative logarithmic returns are stationary.

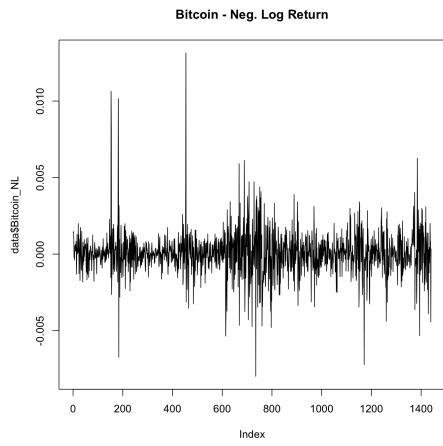


Figure 1. Bitcoin Neg. Log Return

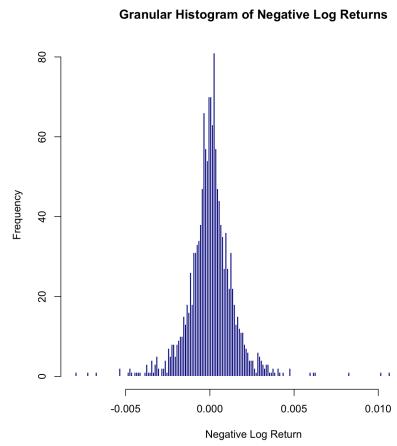


Figure 2. Histogram Neg. Log Return

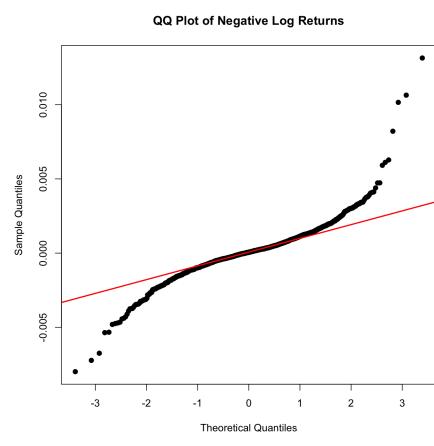
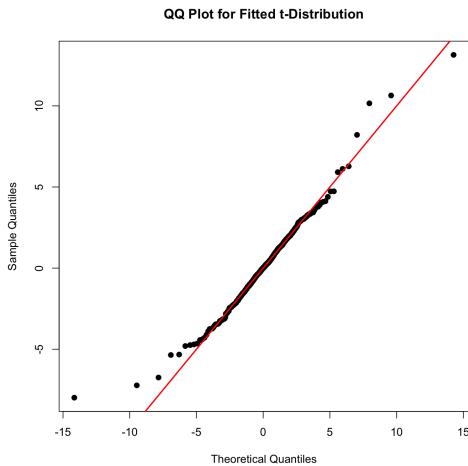
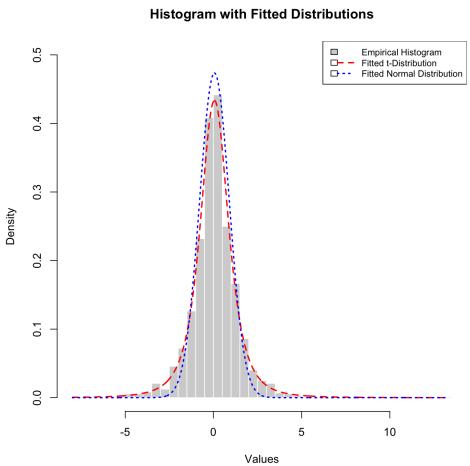


Figure 3. QQplot Neg. Log Return

- c) We observe a certain degree of normality in the histogram (Figure 2), although it displays heavy tails. The QQ plot (Figure 3) confirms this observation: while the points align well with the theoretical line at the centre, they deviate significantly at the extremes, indicating substantial deviations in the tails. The Anderson-Darling (AD) test further confirms that negative returns are not normally distributed. Specifically, since the p-value of the AD test (2.2e-16) is smaller than the common 1% threshold, we reject the null hypothesis that the negative log returns follow a normal distribution.
- d) The student's t-distribution is a probability distribution similar to the normal distribution (bell-shaped) but with heavier tails. It is more likely to produce values far from its mean and, therefore, better suited for modelling data with extreme values or outliers. Visually (Figure 4), we can see that the t-distribution fits the data better than the normal distribution. Although outliers are still present, there is no systematic deviation from the 45-degree line, unlike what was observed with the normal distribution.
- e) The t-distribution is more effective at predicting extreme events than the normal distribution. In Figure 5, this is reflected in an overall better fit. The t-distribution shifts some of the density from the central values to the extremes, allowing it to capture the large value fluctuation observed in the empirical data. In contrast, the normal distribution seemingly fails to capture this variation. In conclusion, the t-distribution fits our data better because it better captures the occurrence of observations far from the mean.

**Figure 4.** QQplot fitted t-Distribution**Figure 5.** Histogram fitted distributions

1.2 Part 2: Financial Time Series, Heteroscedasticity and the Random Walk Hypothesis

a) Based on the ACF plots, negative logarithmic returns are easier to model compared to raw Bitcoin prices. The ACF of raw prices (Figure 27) shows strong autocorrelations across multiple lags, indicating non-stationarity and a high level of dependency, which makes modelling challenging.

In contrast, the ACF of negative logarithmic returns (Figure 28) displays well-defined, near-zero autocorrelations at most lags, suggesting stationarity and behaviour similar to a white noise process. This transformation reduces the impact of extreme values, stabilizes variance, and simplifies the modelling process, making negative logarithmic returns the preferred choice for analysis.

b) The Ljung-Box test allows us to assess whether a group of autocorrelations in a time series is significantly different from zero, which would indicate temporal dependence. For the raw price series, the p-value is extremely low ($< 2.2e-16$), meaning we reject the null hypothesis of no autocorrelation. This indicates significant temporal dependence in the raw data. For the negative logarithmic returns, the p-value is 0.03147, which means that at the classical significance level of 1%, we cannot reject the null hypothesis of no autocorrelation. Therefore, the negative returns do not show significant temporal dependence at this threshold. However, they do at the 5% level, leading to mixed results.

c) We selected an ARIMA(2,0,0) model. We chose $p = 2$ because of the series' exponential decay and a significant peak at lag 2 visible in the ACF graph (Figure 28). Since the negative logarithmic returns are already stationary, d was set to 0. Finally, because the ACF plot only showed minimal spikes over the displayed lags, we set the moving-average component to zero, $q = 0$.

Auto ARIMA proposed ARIMA(2,0,2), indicating that the spikes were stark enough to set $q = 2$. We assessed residual autocorrelation with the Ljung-Box test. For ARIMA(2,0,2), a p-value of 0.5864 confirms our observation of no significant autocorrelation. ARIMA(2,0,0) indicates the same findings ($p = 0.2873$). Both models conclude that the residuals are not temporally correlated.

Residuals from both models are roughly normally distributed. Although ARIMA(2,0,2) performs slightly better in the Ljung-Box test and has fewer significant lags in the ACF (Figures 29 and 30), ARIMA(2,0,0) is simpler. Both models remain appropriate.

d) Both models appear to exhibit minimal autocorrelation in the standardized residuals. However, the GARCH model

with a t-distribution shows better performance when examining the QQ plot (Figures 31 and 32). That is because it handles heavy tails, common for financial time series, better. The Ljung-Box test for both models indicates that there is no significant autocorrelation in the negative log returns, with p-values of 0.3422 for the model assuming normal distribution and 0.3506 for the model assuming t-distribution.

e) Looking solely at the graphs (Figures 33 and 34), there seems to be no difference between the two models. However, the t-distribution continues to outperform the normal distribution, visible in the QQ-plot, where it better captures the heavy tails/extreme values. Additionally, both GARCH-ARIMA models, i.e. normal and t-distribution, capture the absence of autocorrelation of the standardized residuals effectively ($p = 0.3395$ and $p = 0.4062$).

f) ARIMA models are well-suited for stationary series with constant mean and variance but often fail to capture volatility fluctuations, such as volatility clusters. They are, therefore, less ideal for modelling financial data of highly volatile vehicles like Bitcoin. GARCH models are built to handle time-varying volatility, effectively capturing Bitcoin's high and fluctuating volatility. However, they do not control for temporal dependence between past and present returns. The ARIMA-GARCH model combines the strengths of both models, capturing temporal dependence and time-varying volatility in the data. It is, therefore, to be preferred for financial time series modelling. Using a normal distribution, the ARIMA-GARCH model shows improvement over GARCH alone, reflecting its broader applicability.

Only the ARIMA model assumes homoscedasticity. It is, therefore, not suited to model heteroskedastic data like financial time series. The GARCH and ARIMA-GARCH models do not violate this assumption because they consider changes in variance over time.

1.3 Part 3: Dependence Between Time Series

a) The null hypothesis in a correlation test is $corr = 0$, which indicates no relationship. The negative log returns of Bitcoin and Ethereum do not seem to be instantaneously ($lag = 0$) dependent. Although the correlation is slightly negative, it is very weak (-0.002531565). T p-value of 0.9235 confirms that there is no significant correlation between the two series. Therefore, the two series can be considered independent. Finally, by constructing the 95% confidence interval, we observe that 0 is included, which further confirms the above findings.

b) In Figure 36, we observe a strongly significant negative lag, suggesting that the negative log returns of Bitcoin precede those of Ethereum. At a lag of approximately -5, a significant cross-correlation of over 80% is visible, suggesting a lead-lag relationship between Bitcoin and Ethereum.

c) We conducted Granger causality tests for lags of 1, 5, and 10 periods, as is common in financial analysis. The tests were performed in both directions: Bitcoin on Ethereum and Ethereum on Bitcoin. The results show that Bitcoin is a very good predictor of Ethereum, with highly significant p-values at all tested lags (Table 1). In contrast, Ethereum does not seem to predict Bitcoin, regardless of the lag. These findings confirm a lead-lag relationship between Bitcoin and Ethereum.

Lag	BTC → ETH	ETH → BTC
1	0.0131 **	0.2935
5	< 2.2e-16 ***	0.71
10	< 2.2e-16 ***	0.8107

Table 1. Granger Causality Test Results

- d) 1) Since Bitcoin seems to predict Ethereum but not the other way around, we expect that a significant drop in Bitcoin will lead to a similar drop in Ethereum with a lag of about 5min.
 2) In contrast, a sudden drop in Ethereum should not have a notable effect on Bitcoin since the latter appears to evolve independently of Ethereum's movements.

2 Practical 2

2.1 Part 1: Block Maxima Approach

a) As seen in Figure 6, the data is asymmetrically distributed with a strong concentration of zero-values and thin tails. The majority of days experience little to no precipitation, while only a few days record extreme precipitation. The exponential distribution would fit well because it effectively models the observed rapid decline in the value frequency of strictly positive precipitation observations.

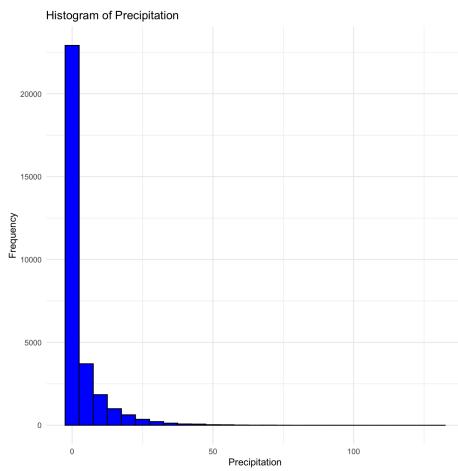


Figure 6. Histogram of precipitation

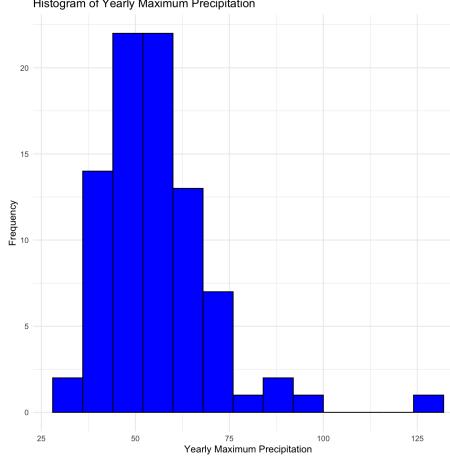


Figure 7. Histogram of Yearly max precipitation

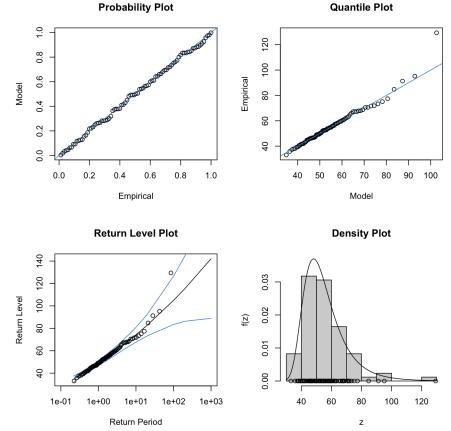


Figure 8. Diagnostic Plot GEV fit

- b) Investigating the yearly maximum values of precipitation (*blocksize = 1year*), the distribution shifts its mass to the right but retains its flat tail. The histogram (Figure 7) indicates that the annual maximum precipitation values are concentrated within a specific range, with a moderate decline in frequency as values increase. The Gumbel GEV distribution appears to be the best choice for this data, as it effectively captures the light-tailed nature and central tendency observed in the data.
- c) This approach is not reasonable, as a linear model fails to capture the magnitude and variability of extreme observations. The fitted values revolve around the mean, starting at 55.8 mm and slightly decreasing to 55.3 mm. The extrapolation follows this logic, with predicted values linearly decreasing from 55.3 to 55.2 mm. The model fails to reflect the significant fluctuation and magnitude of past yearly maximum values, limiting the model's relevance for analyzing and predicting extreme precipitation events.
- d) The model with constant parameters shows lower values of the AIC (672.94) and BIC (680.27) than the model with time-varying location parameters ($AIC = 674.89$ and $BIC = 684.66$). Since lower values indicate better fit, while penalizing model complexity, the constant model should be preferred.
- e) By observing the diagnostic plots (Figure 8), we conclude that the GEV model with constant parameters fits the data well. The points generally follow the expected trends in the probability, quantile, and return level plots, indicating a good fit. There is one outlier that seemingly significantly deviates from the distribution. It does, however, not lower the general appropriateness of the model.
- f) The 10-year return level is estimated at 73.60759 mm (Figure 9). It means that, on average, daily precipitation exceeding a level of 73mm is expected to occur once every 10 years.
- g) The amount of historical values above the requested return levels are visible in Table 2. Unlike the linear model, the

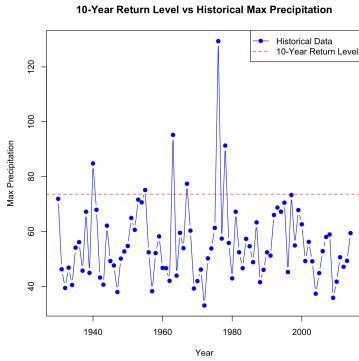


Figure 9. 10-Years return level vs Historical max precipitation

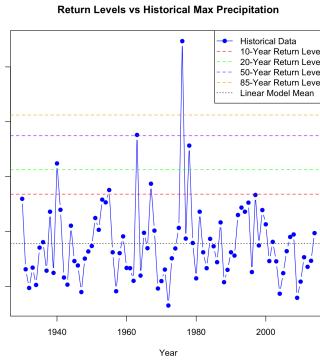


Figure 10. Return level vs Historical max precipitation

Model	Return Level	Values above
10 -Year Return Level:	73.60759	6
20 -Year Return Level:	82.53069	4
50 -Year Return Level:	94.9044	2
85 -Year Return Level:	102.4527	1
Linear model:	55.56	35

Table 2. Historical values above different return levels

GEV model effectively captures the negative relationship between magnitude and occurrence of extreme precipitation events. The linear model does not differentiate for magnitude and occurrence and systematically underestimates the occurrence of extreme events (exceeded 35 times)(Figure 10).

- h) With the adjusted GEV model, observing daily precipitation of 100 mm corresponds to an estimated return period of 71.71 years.
- i) The probability of experiencing more than 150 mm of rain on a single day in a given year is about 0.064%.

2.2 Part 2: Peaks-Over-Threshold (POT) Approach

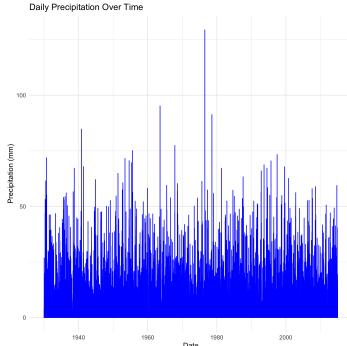


Figure 11. Daily precipitation over time

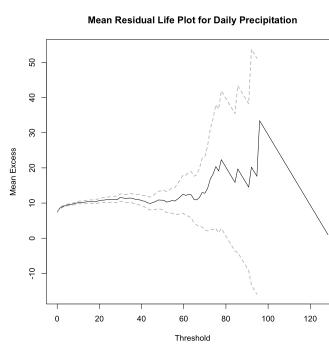


Figure 12. Mean Residual Life plot Daily Precipitation

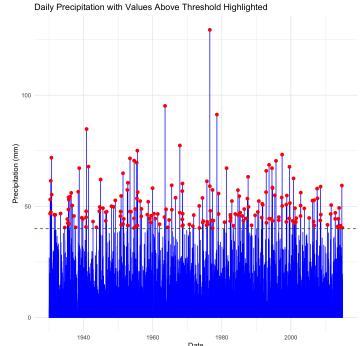


Figure 13. Daily precipitation with Highlighted values above threshold

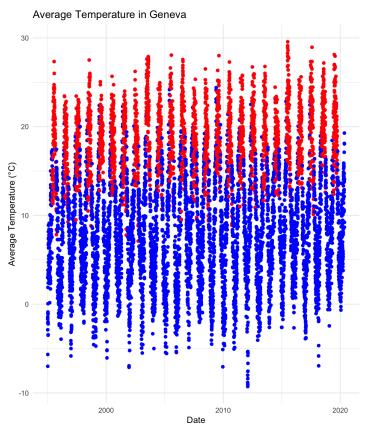
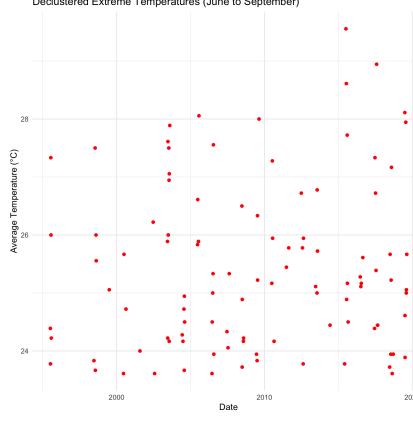
- a) Figure 11
- b) The threshold of 40 mm was chosen from inspection of the Mean Residual Life Plot (Figure 12). Before 40mm, the plot follows a roughly linear trend. After 40 mm, the variance becomes unstable, resulting in significant fluctuations in mean excess. Figure 13 highlights the observations in the data that exceed the 40mm threshold.
- c) The GPD model provides a good overall fit, as shown in the Density Plot and Probability Plot, where the data closely follows the fitted distribution (Figure 38). However, the QQ-Plot and Return Level Plot reveal slight deviations for extreme values (>100 mm), suggesting an underestimation of very large precipitation values. The fit is, therefore, reasonable for frequent to rare events but could be improved for modelling extremely rare events.
- d) Results in Table 3
- e) The return period for a precipitation level of 100 mm is estimated to be approximately 94.81 years.

Table 3. Return Levels for Different Periods

Return Period (Years)	Return Level (Absolute Precipitation)
10	74.36517
20	82.16374
50	92.61074
85	98.73327

- f) The daily probability of exceeding 150 mm of precipitation is extremely low (0.000045%). Over an entire year, the probability of such an event occurring at least once is estimated to be 0.0165%.
- g) The results of the block maxima (0.064%) and POT (0.0165%) approaches roughly align. None of the methods is superior to the other. The block maxima approach only considers the maximum value per block, which can overestimate the occurrence of extreme events if such values are outliers. The POT method accounts for all observations exceeding a certain threshold, which leads to underestimation of extreme events if the chosen threshold is too low. The block maxima approach is simpler to implement, while the POT approach allows for more precise model configuration. We do not have a preference since at correctly chosen block size and threshold level, the approaches should yield similar results.

2.3 Part 3: Clustering and Seasonal Variations

**Figure 14.** Average temperature in Geneva**Figure 15.** Declustered Extreme temperature in summer

a) Result in Figure 14
b) We use the 90th percentile of temperatures to determine the threshold. This means we consider the top 10% of values as extreme. We find an extremal index of 0.2036531. Since it is far from 1 and close to 0, extreme events in temperature do not seem to be isolated events but occur in clusters. The average cluster size is calculated as $1/0.2037$, which equals approximately 4.9 days. Accordingly, extreme temperatures tend to cluster together for about 4 to 5 consecutive days. The probability of another extreme event occurring the day after an extreme event is approximately $1 - 0.2037$, or about 79.63%.

- c) Result in Figure 15
d) The model with declustered data has an AIC of 323.5845 and a BIC of 328.8344, while the model with raw data has an AIC of 886.481 and a BIC of 893.8952. The values indicate that the declustered model provides a better fit to the data, as reflected by its lower AIC and BIC scores.

The 10-year return levels for both models are similar. For the raw data, the 10-year return level is estimated at 30.12°C, while for the declustered data, it is slightly higher at 30.15°C. This proximity in estimates demonstrates that, despite differences in data processing, both models converge toward a consistent evaluation of extreme intensities for a given return period.

3 Practical 3

Over the last couple of years, eggs gained more and more attention. The rise of [world_record_egg](#), surpassing 60 million likes on Instagram, and an increasing number of egg recipes reflect the increase in attention to eggs globally. Especially for small shop owners, analysing the occurrence of extreme demand changes in their egg sales, therefore, becomes of increased importance. For that reason, we picked a dataset of daily egg sales of a local store in Sri Lanka between 1993 and 2020 and analysed its extreme values to derive strategically valuable knowledge for local shop owners.

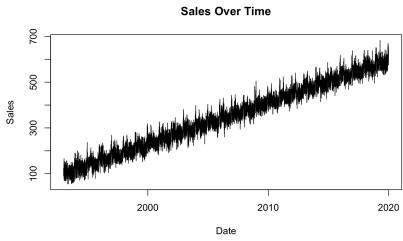


Figure 16. Egg Sales

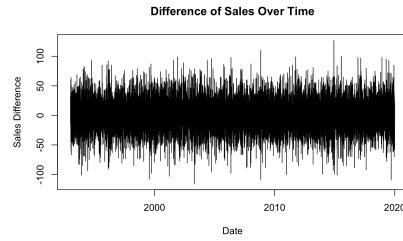


Figure 17. Egg Sales Difference

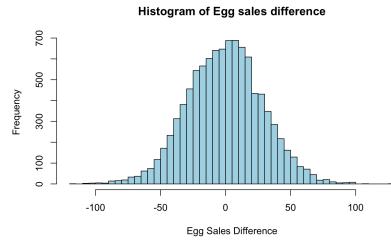


Figure 18. Histogram

3.1 Data

The data was retrieved on the 7th of December from [Kaggle](#) and contains daily egg sales between 1993 and 2020. The raw data can be seen in Figure 16. From visual inspection, the data seems to follow a linear upward trend, confirmed by the Mann-Kendall test ($p \ll 0.05$). We, therefore, decided to analyse the difference in egg sales for two consecutive days, i.e. "egg returns". The final output can be seen in Figure 17. The data is not skewed (-0.009), and stationary according to the ADF test ($p = 0.01$). In order to make forecasts on our data we investigated which distribution would best fit our data. Various distributions were fitted to assess the nature of the data. Given the bell shape of the histogram, visible in Figure 18, a normal and t-distribution were fitted on the data.

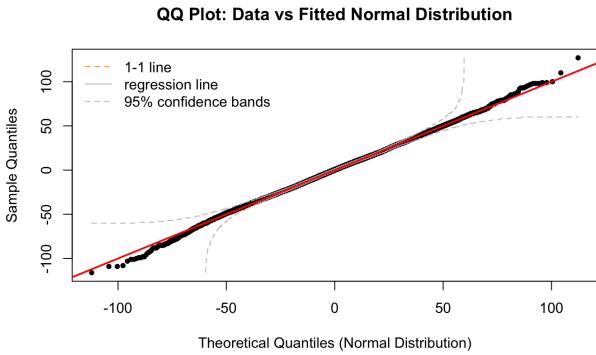


Figure 19. QQ-plot against normal-distribution

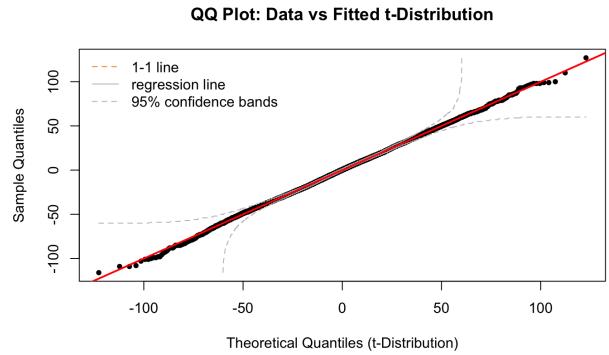


Figure 20. QQ-plot against t-distribution

Investigating the QQ-plots of both, the normal and the t-distribution (Figures 19 and 20), we find that the t-distribution seems to approximate the data slightly better than the normal distribution. Towards the tails, the residuals of the t-distribution seem smaller than those of the normal distribution. This indicates that events far from the mean occur more often than in a normal distribution (heavy-tailed data) and are therefore better approximated by the t-distribution.

3.2 Extreme Value Analysis

In the following three subsections, two extreme value analyses were conducted to analyse the occurrence and magnitude of extreme changes in egg sales. To not exceed the scope of this analysis, we intentionally only reduced the scope to *positive* changes in egg sales.

3.2.1 Block Maxima Approach

For the block maxima approach, we took $blocksize = month$. Accordingly, 9,861 daily observations were grouped into 336 monthly observations from which the maximum change of egg sales between two consecutive days was chosen. The resulting time series is depicted in Figure 21. The maximum value distribution is shown in the histogram in Figure 22 below.



Figure 21. Monthly Maxima over Time

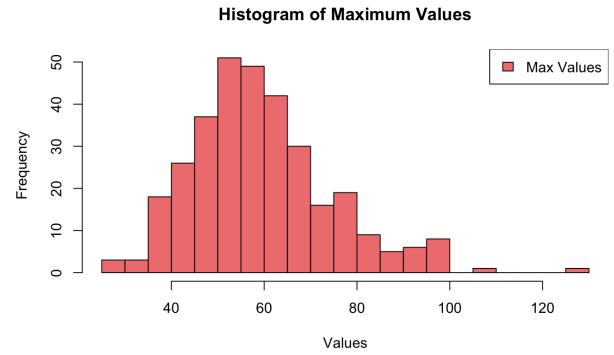


Figure 22. Histogram of Maxima

We chose a constant GEV model over the GEV with time-varying location parameters due to the former's lower BIC ($2664.534 < 2665.849$) and lower residuals in the Q-Q plot. We first want to estimate the return levels of positive egg sales changes within the next years, so our Sri Lankan shop owner knows what he might have to expect in the coming year. The results are listed in Table 4.

Period	Return Level
1 year	82
2 years	90
5 years	99
10 years	105

Table 4. Return Levels for Different Periods

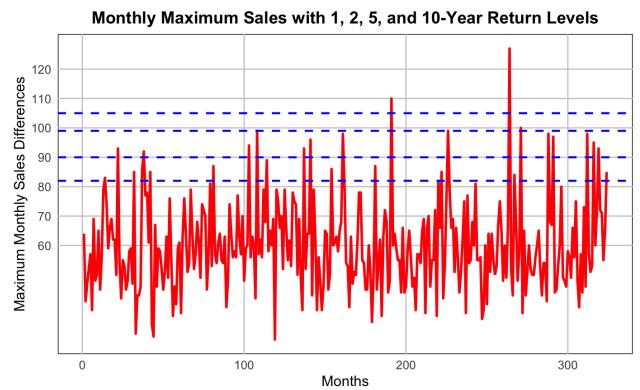


Figure 23. Monthly Extreme Values and Return

For a one-year period, the return level of 82 eggs difference indicates that within the next year, the egg sales of the Sri Lankan shop owner will increase by 82 between two consecutive days at least once. The return level for the 10-year period is 105. Compared to the historical data (visualised including other return levels in Figure 23), this threshold was surpassed

twice in the last 30 years, which is in line with our calculations.

To investigate whether the shop owner should invest in larger egg storage to be able to respond to sudden spikes in demand, the time periods for given return levels also need to be estimated.

Table 5. Return Periods for Different Return Levels with referring Probability to occur in the next year

Return Level	Return Period	Probability
50	1.34 months	99%
75	6.49 months	86.59%
90	2.03 years	39.46%
100	5.47 years	16.82%
120	50.8 years	1.94%

Table 5 shows different return levels and their corresponding return periods. Additionally, for interpretability, the probability of occurrence (over one year) of each return level was calculated.

3.2.2 POT Approach

To enhance the robustness of the block maxima analysis above, the POT method is applied. It is expected to provide a more detailed understanding of extreme variations in positive egg sales differences.

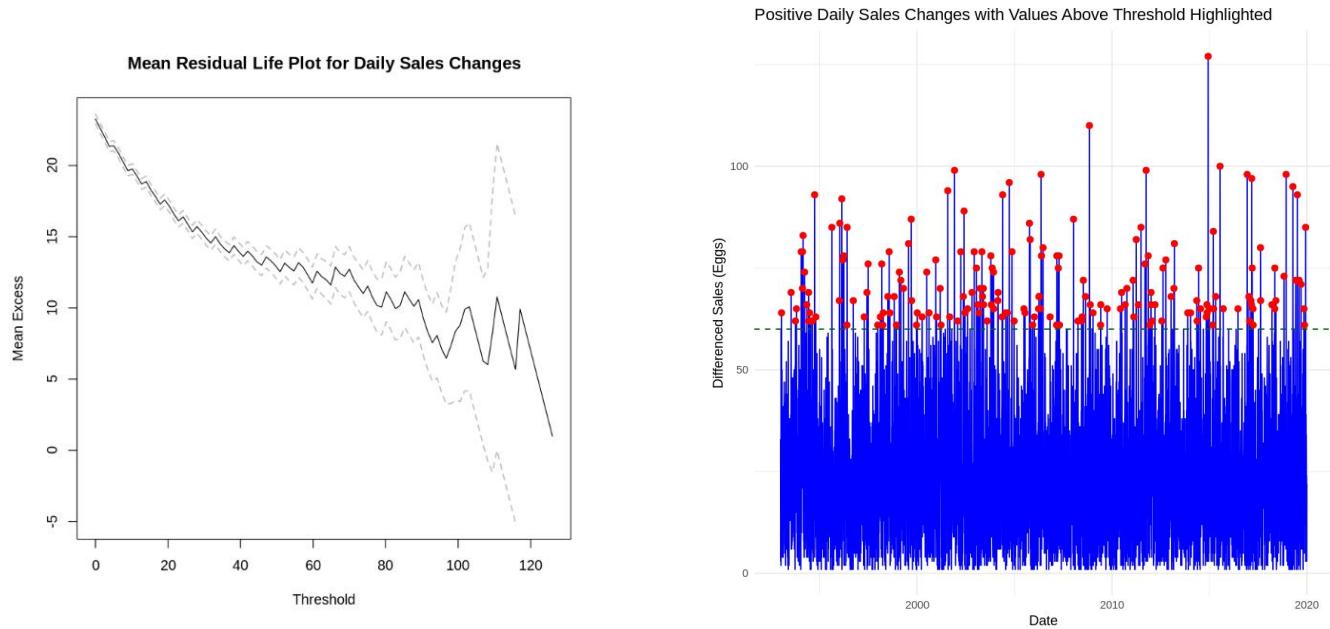


Figure 24. Mean Residual Life Plot for positive Daily Sales Change

Figure 25. Daily positive Sales difference with Values over Threshold highlighted

From visual inspection of Figure 24, a threshold of 60 unit differences has been chosen. Before this point, the mean excess exhibits approximate linearity with increasing threshold levels. Beyond a difference of 60 egg sales, the variance becomes unstable, and the confidence interval explodes. Accordingly, Figure 25 highlights all positive sales changes exceeding the threshold of 60 sales difference of eggs.

The GPD model (Annex Figure 39) demonstrates a generally good fit, particularly in the Probability Plot and Mean Residual Life Plot, where the empirical largely correspond the theoretical observations. However, the Q-Q plot suggests some deviation for high values, indicating a potential underestimation of extremes. While the model is effective in capturing central tendencies, it may require adjustments to better account for the variability in extreme values.

Return Level	Return Period	Probability
50	0.80 months	99%
75	6.05 months	86.32%
90	1.97 years	39.90%
100	5.23 years	17.39%
120	45.9 years	2.16%

Table 6. POT - Return Periods for Different Return Levels with referring Probability to occur in the next year

Period	Return Level
1 year	83
2 years	90
5 years	99
10 years	106

Table 7. POT - Return Levels for Different Periods

Table 6 shows different return periods per return level, including the corresponding probability of occurrence (over one year). The return periods are slightly lower, and the probabilities correspondingly slightly higher, than those obtained using the block maxima approach. Overall, our POT results reinforce the findings of the block maxima method. Table 7 shows the return levels for different periods. Results here, too, are validated by their similarity with the results of the block maxima approach.

The POT and Block Maxima approaches yield consistent results in terms of return levels and return periods. The similarity of our results thereby reinforces the confidence in the accuracy of our estimations.

3.2.3 Clustering and Holidays

We initially examined the data to check for clusters of extreme values without segmenting the data by specific months or holidays. The results were inconclusive, with an extremal index of 0.9288 for daily sales, suggesting no apparent clustering of extreme values. Suspecting that the dispersion of extreme values might obscure potential patterns, we grouped the daily sales data into weekly and monthly intervals, expecting the aggregation to reveal potential interpretational relevance. However, the extremal indices for both cases were 1, too, indicating no clustering of extremes.

Our final attempt to identify extreme value clusters was to focus specifically on April. It is a culturally significant month in Sri Lanka, celebrated with the Sinhala and Tamil New Year and, with minor relevance, Eastern. It is a time of festivities, family gatherings, and traditional culinary practices, where eggs are a key ingredient to festive dishes such as *kokis* (a crispy rice-flour delicacy), *butter cake* and various other sweets. These celebrations are expected to lead to heightened household consumption and a consequential increase in market activity for staple items. We isolated the days of April from the dataset and recalculated the extremal index. The result was 0.9539, indicating no significant clustering of extreme values. Despite this, we proceeded to decluster the data to observe if any patterns emerged. The declustered data yielded an extremal index of 0.9534 — an insignificant difference, reinforcing the absence of clustered extreme sales. Our findings make sense because the holiday month is a seasonal event that should theoretically not lead to unexpected egg-demand spikes. Such cyclical consumption patterns are foreseeable by a shop owner and are therefore corrected for in our analysis when the time series was made stationary. It is, therefore, not surprising that we do not identify clusters during cyclical holiday months.

In conclusion, our analysis indicates that there are no clusters of extreme sales within the dataset, suggesting that extreme egg sales occur largely isolated in Sri Lanka, also during culturally significant months like April.

Annexes

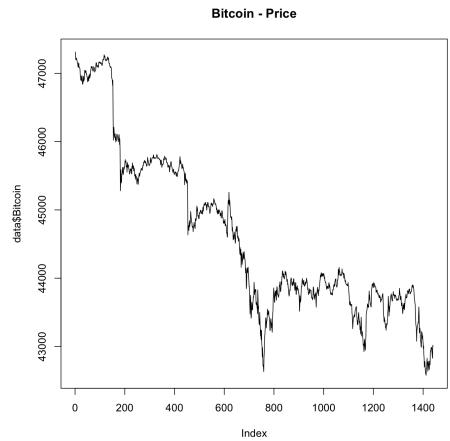


Figure 26. Bitcoin Prices

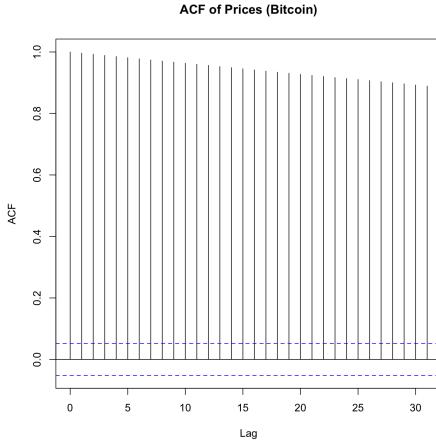


Figure 27. ACF Prices

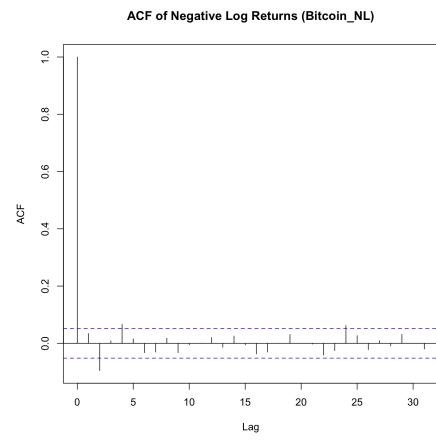


Figure 28. ACF Neg. Log Return

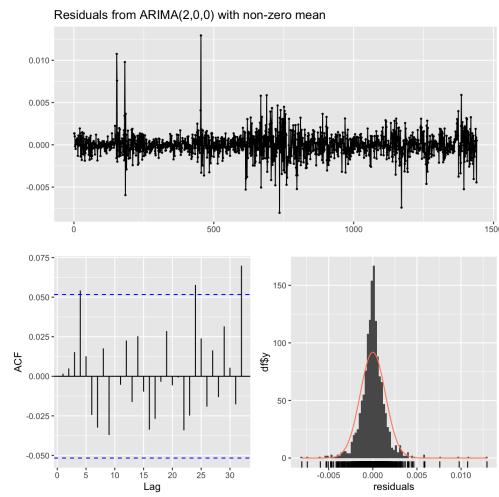


Figure 29. Residuals ARIMA(2,0,0)

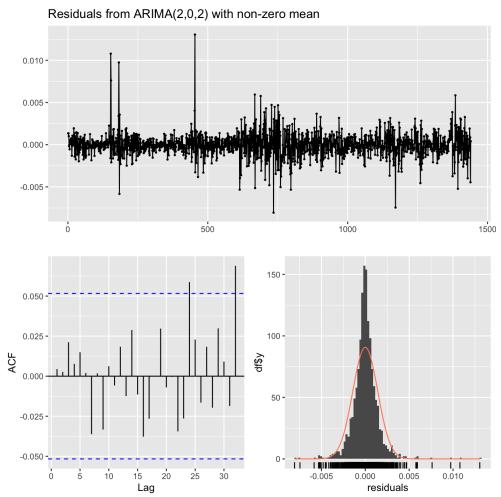
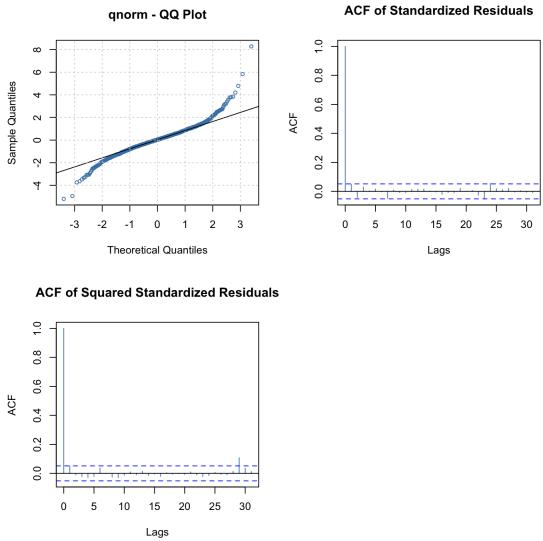
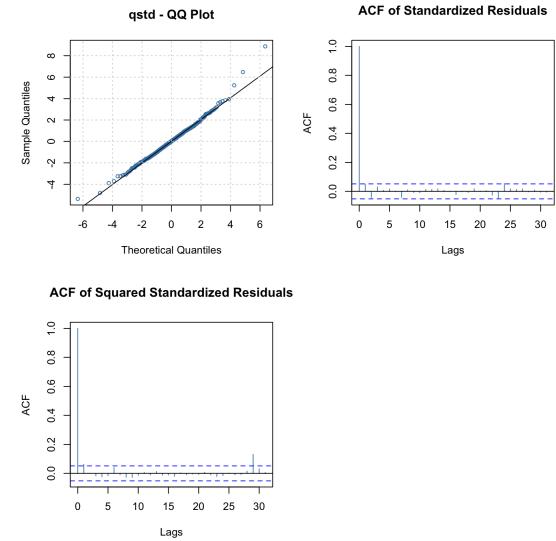
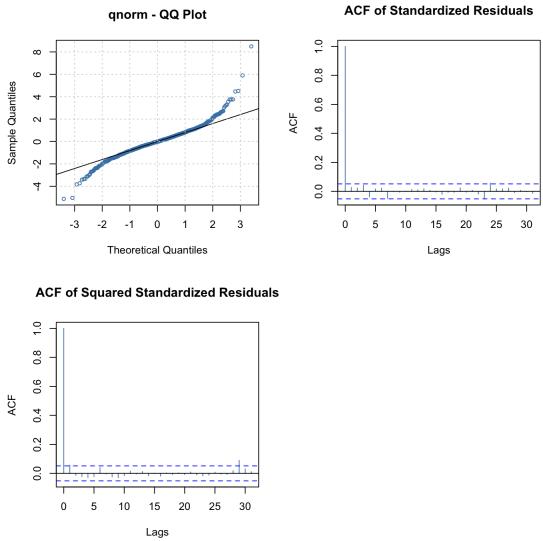
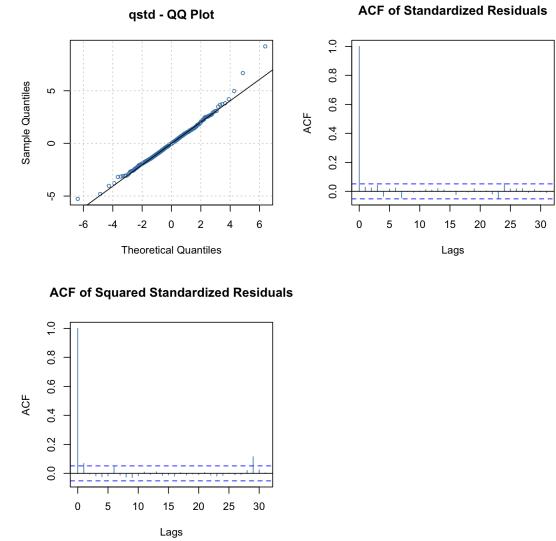
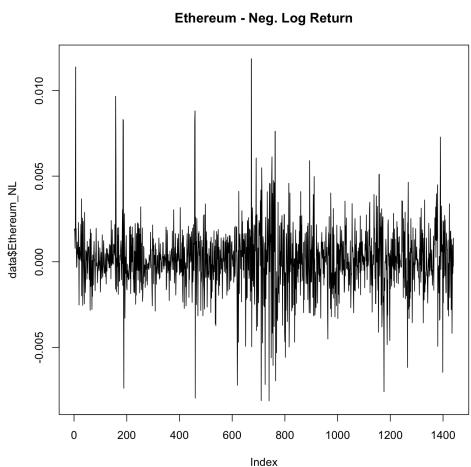
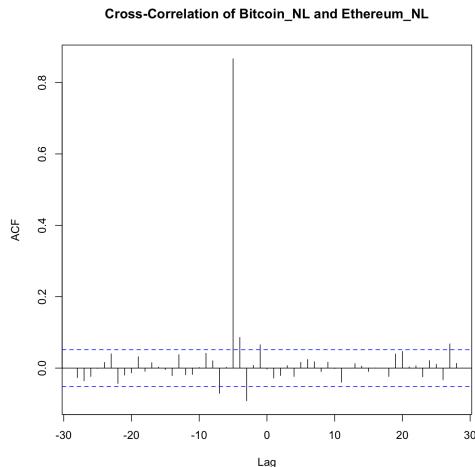
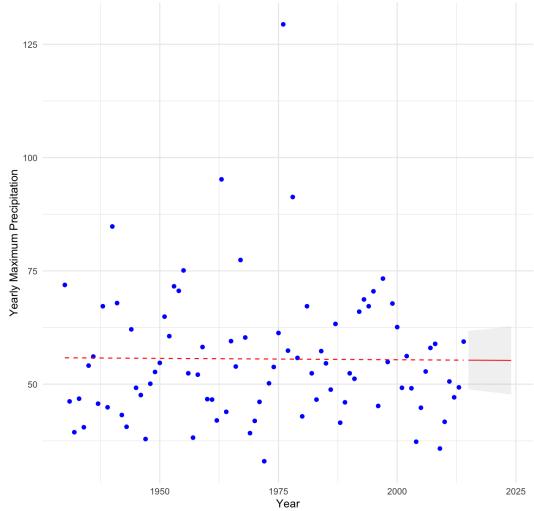
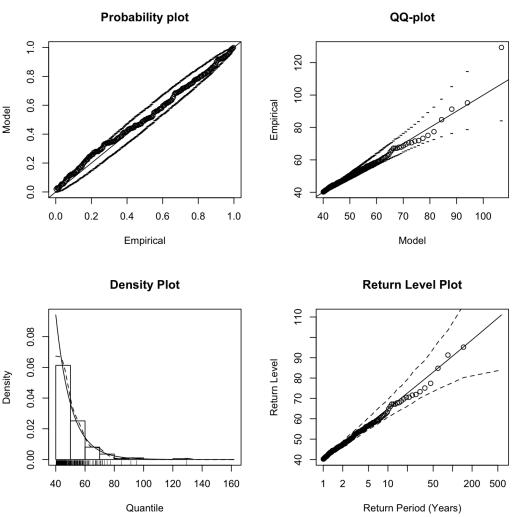
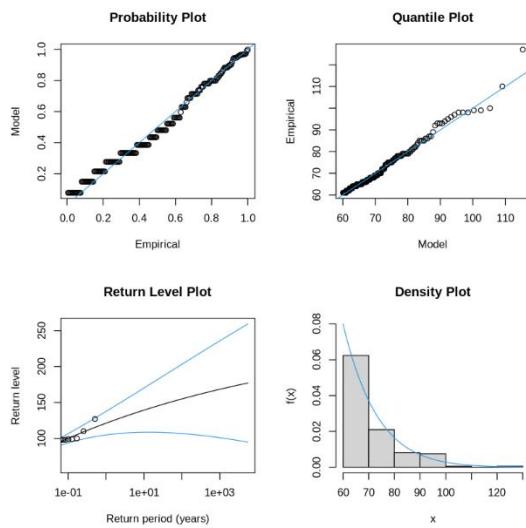


Figure 30. Residuals ARIMA(2,0,2)

**Figure 31.** GARCH Normal distribution**Figure 32.** GARCH T-distribution**Figure 33.** GARCH-ARIMA Normal distribution**Figure 34.** GARCH-ARIMA T-distribution**Figure 35.** ETH Neg. Log Return**Figure 36.** Cross-Correlation Bitcoin Ethereum

Yearly Maximum Precipitation with Linear Model and Predictions

**Figure 37.** Yearly max Precipitation - Linear Model and Predictions**Figure 38.** Diagnostic plot GPD**Figure 39.** Diagnostic plot POT