SA2

Mercado, C & Sinocruz, A.

2025-05-17

2. Test using Shapiro-Wilk normality test the Ethereum returns for trading data every five minutes, from August 7, 2015 to April 15, 2025.

```
library(dplyr)
library(readr)
library(lubridate)
library(ggplot2)
```

dataset

```
# Load first dataset (2015 to 2020)
df1 <- read_csv("C:/Users/CONSUELO B. MERCADO/Downloads/sa2/ETHUSDT_2015_to_2020.csv", show_col_types =
  rename(date = Date, close = Close)

# Load second dataset (2020 to 2025)
df2 <- read_csv("C:/Users/CONSUELO B. MERCADO/Downloads/sa2/ETHUSDT_2020_to_2025.csv", show_col_types =
  rename(date = date, close = close)

# Combine and clean
combined_df <- bind_rows(df1, df2) %>%
  filter(!is.na(date)) %>%   # remove rows with missing dates
  arrange(date)

# Calculate log returns
combined_df <- combined_df %>%
  mutate(log_return = log(close / lag(close))) %>%
  filter(!is.na(log_return))

# Sample 5000 or fewer log returns
set.seed(42)
sample_size <- min(5000, nrow(combined_df))
sample_returns <- sample(combined_df$log_return, sample_size)

# Shapiro-Wilk Normality Test
shapiro_test <- shapiro.test(sample_returns)
```

```r
# Print results
cat("  Shapiro-Wilk Normality Test on ETH/USDT 2015-2025 5-min Returns\n")
```

```
##   Shapiro-Wilk Normality Test on ETH/USDT 2015-2025 5-min Returns
```

```r
cat(sprintf("Test Statistic: %.5f\n", shapiro_test$statistic))
```

```
## Test Statistic: 0.53694
```

```r
cat(sprintf("P-value: %.5e\n", shapiro_test$p.value))
```

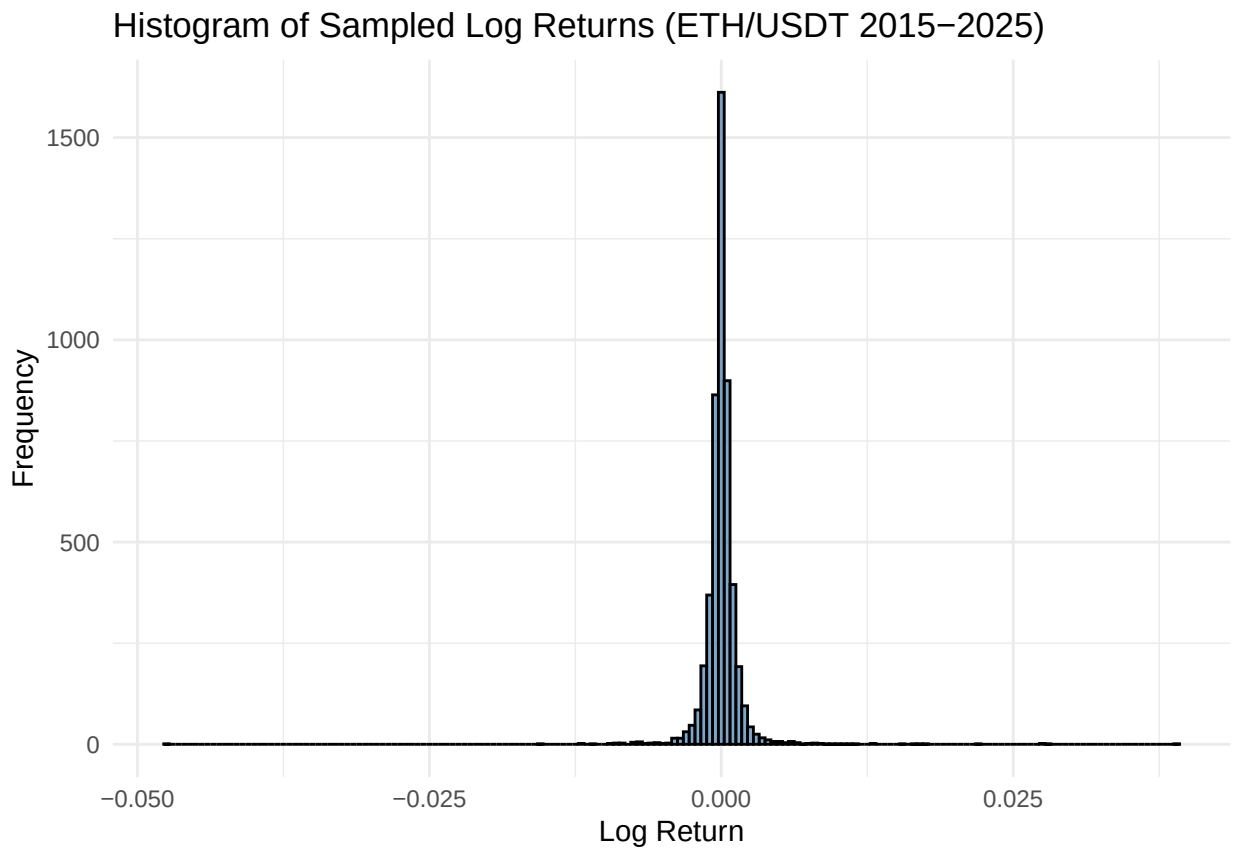
```
## P-value: 1.69021e-78
```

```r
if (shapiro_test$p.value > 0.05) {
  cat("  Returns appear normally distributed (fail to reject H ).\n")
} else {
  cat("  Returns are not normally distributed (reject H ).\n")
}
```

```
##   Returns are not normally distributed (reject H ).
```

Histogram

```r
# Plot histogram of sampled log returns
ggplot(data.frame(log_return = sample_returns), aes(x = log_return)) +
  geom_histogram(binwidth = 0.0005, fill = "steelblue", color = "black", alpha = 0.7) +
  labs(
    title = "Histogram of Sampled Log Returns (ETH/USDT 2015-2025)",
    x = "Log Return",
    y = "Frequency"
  ) +
  theme_minimal()
```



Results Interpretation

The normality of Ethereum 5-minute log returns from August 7, 2015 to April 15, 2025 was assessed using two statistical tests:

Shapiro-Wilk Test (on a random sample of 5,000 returns):

- Test Statistic: 0.53694
- P-value: 1.69021e-78
- **Interpretation**: The very low p-value indicates a strong rejection of the null hypothesis. This means the sampled Ethereum returns do not follow a normal distribution.