Methodology for Commodity Returns Analysis

Manuel J. Cerezo

October 27, 2024

1 Introduction

This document provides an overview of the methodology used in the **Commodity Returns Analysis** project. The project analyzes historical log returns of various commodities, exploring distribution characteristics, normality, and underlying patterns through advanced statistical techniques and visualization. The code is a practical application of the first chapters of *"Statistics and Data Analysis for Financial Engineering: with R examples"* by David Ruppert and David S. Matteson.

2 Methodology

2.1 Data Collection

Commodity price data is sourced from Yahoo Finance using the 'pandas datareader' and 'yfinance' libraries. The data retrieval function accesses daily closing prices for each commodity, covering a specified date range.

The API request is structured as follows:

```
from pandas_datareader import data as pdr
import yfinance as yfin
yfin.pdr_override()
data = pdr.get_data_yahoo("COMMODITY_SYMBOL", start="YYYY-MM-DD", end="YYYY-MM-DD")
```

2.2 Log Return Calculation

Log returns are used to analyze the relative change in prices, defined as:

$$\text{Log Return}_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1}$$

where P_t represents the closing price on day t.

2.3 Normality Tests

To assess the normality of the log returns, we use the following methods:

- **QQ Plot**: A Quantile-Quantile (QQ) plot compares the quantiles of the log returns to those of a normal distribution. Deviations from the reference line indicate departures from normality.
- **Shapiro-Wilk Test**: This statistical test evaluates the hypothesis that the log returns follow a normal distribution. A low p-value indicates a rejection of normality.

2.4 T-Plot

The T-plot visualizes the temporal dynamics of log returns, highlighting periods of high volatility and identifying potential trends over time. This plot is helpful in detecting clusters of volatility, which are typical in financial time series data.

2.5 Kernel Density Estimation (KDE)

To estimate the probability density function of the log returns, we use Kernel Density Estimation (KDE). The KDE provides a smooth, non-parametric estimate of the distribution:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{2}$$

where K is the kernel function (typically Gaussian) and h is the bandwidth parameter that controls the smoothness of the density estimate.

3 Assumptions

- Log returns are assumed to be independently distributed, though normality is tested through statistical methods.
- Kernel Density Estimation provides a reliable approximation of the underlying distribution.

4 Conclusion

The **Commodity Returns Analysis** project uses log returns, normality tests, T-plots, and KDE to examine the statistical properties of commodity returns. These tools provide insights into the behavior and distribution of returns, which are useful for market analysis and decision-making.