

# 学习笔记：Python for Finance Cookbook, 2nd

Stephen CUI

2023-11-11



# 目录

<b>1</b>	<b>Data Preprocessing</b>	<b>1</b>
1.1	Changing the frequency of time series data . . . . .	1
1.2	Different ways of imputing missing data . . . . .	1
1.3	Different ways of aggregating trade data . . . . .	1
<b>2</b>	<b>Exploring Financial Time Series Data</b>	<b>3</b>
2.1	Outlier detection using rolling statistics . . . . .	3
2.2	Outlier detection with the Hampel filter . . . . .	3
2.3	Detecting changepoints in time series . . . . .	4
2.4	检验资产收益的程式化事实 . . . . .	4
2.4.1	事实 1: Non-Gaussian distribution of returns . . . . .	4
2.4.2	事实 2: 波动性聚类 . . . . .	4
<b>3</b>	<b>回测交易策略</b>	<b>5</b>



# Chapter 1

## Data Preprocessing

### 1.1 Changing the frequency of time series data

The formula for realized volatility is as follows:

$$RV = \sqrt{\sum_{i=1}^T r_i^2} \quad (1.1)$$

Realized volatility is frequently used for calculating the daily volatility using intraday returns.

### 1.2 Different ways of imputing missing data

Two of the simplest approaches to imputing missing time series data are:

- Backward filling—fill the missing value with the next known value
- Forward filling—fill the missing value with the previous known value

### 1.3 Different ways of aggregating trade data

The term **bars** refers to a data representation that contains basic information about the price movements of any financial asset.

冰山订单是大订单被分成较小的限价订单以隐藏实际订单数量。它们被称为“冰山订单”，因为可见的订单只是“冰山一角”，而大量的限价订单正在等待，准备下单。

Ideally, they would want to have a bar representation in which each bar contains the same amount of information. Some of the alternatives they are using include:

- **Tick bars**—named after the fact that transactions/trades in financial markets are often referred to as ticks. For this kind of aggregation, we sample an OHLCV bar every time a predefined number of transactions occurs.
- **Volume bars**—we sample a bar every time a predefined volume (measured in any unit, for example, shares, coins, etc.) is exchanged.
- **Dollar bars**—we sample a bar every time a predefined dollar amount is exchanged. Naturally, we can use any other currency of choice.

“Tick bars” 在金融领域通常指的是一种按照市场价格变动（ticks）来形成的交易条形图或时间序列。这种图表以市场价格的变动为基础，而不是固定的时间间隔。当市场价格变动达到一定的数量（例如，每次价格变动一个tick），就形成一个新的交易条。

这种方法的优势在于，它可以更好地反映市场的活动和波动，而不受时间因素的影响。相比于固定时间间隔的图表，tick bars 可以更好地捕捉市场的高活跃性时段，提供更精确的信息，尤其对于高频交易者和对市场活动敏感的策略来说更为有用。

Each of these forms of aggregations has its strengths and weaknesses that we should be aware of.

Tick bars offer a better way of tracking the actual activity in the market, together with the volatility. However, a potential issue arises out of the fact that one trade can contain any number of units of a certain asset. So, a buy order of a single share is treated equally to an order of 10,000 shares.

Volume bars are an attempt at overcoming this problem. However, they come with an issue of their own. They do not correctly reflect situations in which asset prices change significantly or when stock splits happen. This makes them unreliable for comparison between periods affected by such situations.

That is where the third type of bar comes into play—the dollar bars. It is often considered the most robust way of aggregating price data. Firstly, the dollar bars help bridge the gap with price volatility, which is especially important for highly volatile markets such as cryptocurrencies. Then, sampling by dollars is helpful to preserve the consistency of information. The second reason is that dollar bars are resistant to the outstanding amount of the security, so they are not affected by actions such as stock splits, corporate buybacks, issuance of new shares, and so on.

## Chapter 2

# Exploring Financial Time Series Data

In this chapter, we will cover the following recipes:

- Outlier detection using rolling statistics
- Outlier detection with the Hampel filter
- Detecting changepoints in time series
- Detecting trends in time series
- Detecting patterns in a time series using the Hurst exponent
- Investigating stylized facts of asset returns

### 2.1 Outlier detection using rolling statistics

### 2.2 Outlier detection with the Hampel filter

We will cover one more algorithm used for outlier detection in time series—the Hampel filter. Its goal is to identify and potentially replace outliers in a given series. It uses a centered sliding window of size  $2x$  (given  $x$  observations before/after) to go over the entire series.

For each of the sliding windows, the algorithm calculates the median and the median absolute deviation (a form of a standard deviation).

For the median absolute deviation to be a consistent estimator for the standard deviation, we have to multiply it by a constant scaling factor  $k$ , which is dependent on the distribution. For Gaussian, it is approximately 1.4826.

## 2.3 Detecting changepoints in time series

A **changepoint** can be defined as a point in time when the probability distribution of a process or time series changes, for example, when there is a change to the mean in the series.

## 2.4 检验资产收益的程式化事实

### 2.4.1 事实 1: Non-Gaussian distribution of returns

### 2.4.2 事实 2: 波动性聚类

波动性聚类是一种模式，其中价格的大变化往往会伴随着大的变化（波动性较高的时期），而价格的小变化之后往往会出现小的变化（波动性较低的时期）。

### 2.4.3 事实 3: 收益率缺乏自相关性

### 2.4.4 事实 4: 平方/绝对回报的自相关性较小且递减

While we expect no autocorrelation in the return series, it was empirically proven that we can observe small and slowly decaying autocorrelation (also referred to as persistence) in simple nonlinear functions of the returns, such as absolute or squared returns. This observation is connected to the phenomenon we have already investigated, that is, volatility clustering.



## Chapter 3

# 回测交易策略