

Thársis T. P. Souza

# **OPEN QUANT LIVE BOOK**

**A PRACTICAL, HANDS-ON AND OPEN  
APPROACH TO QUANTITATIVE FINANCE  
ANALYSIS**



# The Open Quant Live Book

Thársis T. P. Souza

2018-12-27



# Contents

<b>Preface</b>	<b>5</b>
Working Content . . . . .	5
 <b>I The Basics</b>	 <b>7</b>
<b>1 I/O</b>	<b>9</b>
1.1 Data Sources . . . . .	9
 <b>2 Stylized Facts</b>	 <b>13</b>
2.1 Introduction . . . . .	13
2.2 Distribution of Returns . . . . .	13
2.3 Volatility . . . . .	14
2.4 Correlation . . . . .	14
 <b>3 Correlation &amp; Causation</b>	 <b>17</b>
 <b>II Algo Trading</b>	 <b>19</b>
 <b>4 Limit Order</b>	 <b>21</b>

<b>III</b>	<b>Portfolio Optimization</b>	<b>23</b>
<b>IV</b>	<b>Machine Learning</b>	<b>25</b>
<b>V</b>	<b>Econophysics</b>	<b>27</b>
<b>5</b>	<b>Entropy</b>	<b>29</b>
<b>6</b>	<b>Transfer Entropy</b>	<b>31</b>
<b>7</b>	<b>Financial Networks</b>	<b>33</b>

# Preface

## Working Content

1. The Basics
  - I/O (25%)
  - Stylized Facts (25%)
  - Correlation & Causation
2. Algo Trading
  - Investment Process
  - Backtesting
  - Factor Investing
  - Limit Order
3. Portfolio Optimization
4. Machine Learning
  - Intro
  - AutoML
  - Hierarchical Risk Parity
5. Econophysics
  - Entropy, Efficiency and Coupling
  - Transfer Entropy, Information Transfer and Causality
  - Financial Networks, Taxonomy and Core-Periphery Structure

Thársis T. P. Souza New York, NY



# Part I

## The Basics



# Chapter 1

## I/O

### 1.1 Data Sources

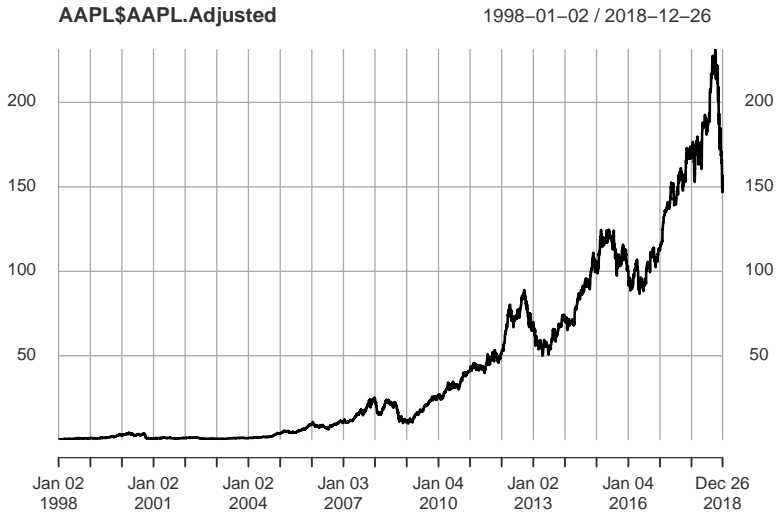
#### 1.1.1 Alpha Vantage

Alpha Vantage offers free access to pricing data including:

- Stock Time Series Data;
- Physical and Digital/Crypto Currencies (e.g., Bitcoin);
- Technical Indicators and
- Sector Performances.

The data are available in JSON and CSV format via REST APIs. The **quantmod** and the **alphavantage** R packages offer a lightweight R interface to the Alpha Vantage API. For instance, daily stock prices can be obtained with the `quantmod::getSymbols` function as follows:

```
getSymbols(Symbols = "AAPL", src = "av", output.size = "full",  
  adjusted = TRUE, api.key = "your API key")  
  
plot(AAPL$AAPL.Adjusted)
```



We called the `quantmod::getSymbols` function with the following arguments:

- `Symbols='AAPL'` defines a character vector specifying the names of each symbol to be loaded, here specified by the symbol of the company Apple Inc.;
- `src="av"` specifies the sourcing method, here defined with the value corresponding to Alpha Vantage;
- `output.size="full"`, strings `compact` and `full` are accepted with the following specifications: `compact` returns only the latest 100 data points; `full` returns the full-length time series of up to 20 years of historical data;
- `adjusted=TRUE`, defines boolean variable to include a column of closing prices adjusted for dividends and splits;
- `api.key`, specifies your Alpha Vantage API key.

**1.1.2 IEX**

**1.1.3 Quandl**



# Chapter 2

## Stylized Facts

### 2.1 Introduction

### 2.2 Distribution of Returns

#### 2.2.1 Fat Tails

A distribuição de retornos financeiros apresenta leptokurtose. A ocorrência de eventos extremos é mais provável comparado com uma distribuição normal, i.e., as caudas da distribuição empírica de retornos são mais “pesadas” comparadas com as caudas esperadas supondo uma distribuição normal de probabilidade.

#### 2.2.2 Skewness

A distribuição empírica de retornos é distorcida para esquerda. Retornos negativos são mais prováveis que retornos positivos.

## 2.3 Volatility

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2} \quad (2.1)$$

### 2.3.1 Time-invariance

A volatilidade de retornos financeiros não é constante ao longo do tempo.

### 2.3.2 Volatility Clustering

Eventos extremos são observados próximos um dos outros.

### 2.3.3 Correlation with Trading Volume

O volume de negociação de um ativo tem correlação significativa com a volatilidade do mesmo.

## 2.4 Correlation

$$\rho = \frac{\sum_{t=1}^T (r_t - \hat{r}_t)(s_t - \hat{s}_t)}{\sqrt{\sum_{t=1}^T (r_t - \hat{r}_t)^2} \sqrt{\sum_{t=1}^T (s_t - \hat{s}_t)^2}}, \quad (2.2)$$

onde  $\hat{r}_t$  e  $\hat{s}_t$  são a média amostral de  $r_t$  e  $s_t$ , respectivamente.

### 2.4.1 Time-invariance

A correlação entre duas series temporais de retornos financeiros não é constante ao longo do tempo.



### 2.4.2 Auto-correlation

Retornos financeiros apresentam baixa autocorrelação (linear), exceto em escalas de tempo muito baixas, e.g., minutos, onde há presença de efeitos de microstructura. Por outro lado, a função de autocorrelação do valor absoluto de retornos financeiros decai lentamente com o tempo.

A correlação contemporânea é maior do que a correlação cruzada.



## Chapter 3

# Correlation & Causation



## Part II

# Algo Trading



## Chapter 4

# Limit Order





## Part III

# Portfolio Optimization



## Part IV

# Machine Learning



**Part V**

**Econophysics**



## Chapter 5

# Entropy





## Chapter 6

# Transfer Entropy



## Chapter 7

# Financial Networks