Data-Driven Finance

Min-Yuh Day, Ph.D, Associate Professor
Institute of Information Management, National Taipei University

https://web.ntpu.edu.tw/~myday

1111AIFQA06
MBA, IM, NTPU (M6132) (Fall 2022)
Tue 2, 3, 4 (9:10-12:00) (B8F40)
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Subject/Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2022/09/13</td>
<td>Introduction to Artificial Intelligence in Finance and Quantitative Analysis</td>
</tr>
<tr>
<td>2</td>
<td>2022/09/20</td>
<td>AI in FinTech: Metaverse, Web3, DeFi, NFT, Financial Services Innovation and Applications</td>
</tr>
<tr>
<td>3</td>
<td>2022/09/27</td>
<td>Investing Psychology and Behavioral Finance</td>
</tr>
<tr>
<td>4</td>
<td>2022/10/04</td>
<td>Event Studies in Finance</td>
</tr>
<tr>
<td>5</td>
<td>2022/10/11</td>
<td>Case Study on AI in Finance and Quantitative Analysis I</td>
</tr>
<tr>
<td>6</td>
<td>2022/10/18</td>
<td>Finance Theory</td>
</tr>
<tr>
<td>Week</td>
<td>Date</td>
<td>Subject/Topics</td>
</tr>
<tr>
<td>------</td>
<td>------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>7</td>
<td>2022/10/25</td>
<td>Data-Driven Finance</td>
</tr>
<tr>
<td>8</td>
<td>2022/11/01</td>
<td>Midterm Project Report</td>
</tr>
<tr>
<td>9</td>
<td>2022/11/08</td>
<td>Financial Econometrics</td>
</tr>
<tr>
<td>10</td>
<td>2022/11/15</td>
<td>AI-First Finance</td>
</tr>
<tr>
<td>11</td>
<td>2022/11/22</td>
<td>Industry Practices of AI in Finance and Quantitative Analysis</td>
</tr>
<tr>
<td>12</td>
<td>2022/11/29</td>
<td>Case Study on AI in Finance and Quantitative Analysis II</td>
</tr>
</tbody>
</table>
# Syllabus

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Subject/Topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>2022/12/06</td>
<td>Deep Learning in Finance; Reinforcement Learning in Finance</td>
</tr>
<tr>
<td>14</td>
<td>2022/12/13</td>
<td>Algorithmic Trading; Risk Management; Trading Bot and Event-Based Backtesting</td>
</tr>
<tr>
<td>15</td>
<td>2022/12/20</td>
<td>Final Project Report I</td>
</tr>
<tr>
<td>16</td>
<td>2022/12/27</td>
<td>Final Project Report II</td>
</tr>
<tr>
<td>17</td>
<td>2023/01/03</td>
<td>Self-learning</td>
</tr>
<tr>
<td>18</td>
<td>2023/01/10</td>
<td>Self-learning</td>
</tr>
</tbody>
</table>
Data-Driven Finance
Data-Driven Finance

- Scientific Method
- Financial Econometrics and Regression
- Data Availability
- Normative Theories Revisited
- Debunking Central Assumptions in Finance

Data-driven finance

- Financial context *(theory, model, application)* that is primarily driven by and based on insights gained from data.

Data-driven finance
Robin Wigglesworth (2019)

• Nowadays, analysts sift through non-traditional information such as satellite imagery and credit card data, or use artificial intelligence techniques such as machine learning and natural language processing to glean fresh insights from traditional sources such as economic data and earnings-call transcripts.

Scientific Method

• Generally accepted principles that should guide scientific effort

• The scientific method is an empirical method of acquiring knowledge that has characterized the development of science

• It involves careful observation, applying rigorous skepticism about what is observed, given that cognitive assumptions can distort how one interprets the observation.

Scientific Method

• It involves **formulating hypotheses**, via **induction**, based on such **observations**; experimental and measurement-based testing of **deductions** drawn from the hypotheses; and **refinement** (or elimination) of the hypotheses based on the **experimental findings**.
Normative Finance and Scientific Method

• Normative financial theories mostly rely on assumptions and axioms in combination with deduction as the major analytical method to arrive at their central results.

  • Expected utility theory (EUT) assumes that agents have the same utility function no matter what state of the world unfolds and that they maximize expected utility under conditions of uncertainty.

  • Mean-variance portfolio (MVP) theory describes how investors should invest under conditions of uncertainty assuming that only the expected return and the expected volatility of a portfolio over one period count.

Normative Finance and Scientific Method

• The capital asset pricing model (CAPM) assumes that only the nondiversifiable market risk explains the expected return and the expected volatility of a stock over one period.

• Arbitrage pricing theory (APT) assumes that a number of identifiable risk factors explains the expected return and the expected volatility of a stock over time; admittedly, compared to the other theories, the formulation of APT is rather broad and allows for wide-ranging interpretations.

Financial Econometrics and Regression

• [Financial] econometrics is the quantitative application of statistical and mathematical models using [financial] data to develop financial theories or test existing hypotheses in finance and to forecast future trends from historical data.

• It subjects real-world [financial] data to statistical trials and then compares and contrasts the results against the [financial] theory or theories being tested.

Financial Econometrics and Regression

• One of the major tools in financial econometrics is regression, in both its univariate and multivariate forms
• Regression is also a central tool in statistical learning in general

Data Availability

• Types of (financial) data
  • Financial econometrics is driven by statistical methods, such as regression, and the availability of financial data
  • Theoretical and empirical financial research was mainly driven by relatively small data sets and was mostly comprised of end-of-day (EOD) data
  • Types of financial and other data available in ever increasing granularity, quantity, and velocity.

• Quality and quantity via programmatic APIs
  • Finance professionals have relied on data terminals from Refinitiv or Bloomberg
  • Major breakthrough in data-driven finance via programmatic APIs

# Relevant types of financial data

<table>
<thead>
<tr>
<th>Time</th>
<th>Structured data</th>
<th>Unstructured data</th>
<th>Alternative data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Historical</td>
<td>Prices, fundamentals</td>
<td>News, texts</td>
<td>Web, social media, satellites</td>
</tr>
<tr>
<td>Streaming</td>
<td>Prices, volumes</td>
<td>News, filings</td>
<td>Web, social media, satellites, Internet of Things</td>
</tr>
</tbody>
</table>
Yahoo Finance World Indices

https://finance.yahoo.com/world-indices/

World Indices

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Name</th>
<th>Last Price</th>
<th>Change</th>
<th>% Change</th>
<th>Volume</th>
<th>Intraday High/Low</th>
<th>52 Week Range</th>
<th>Day Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>^GSPC</td>
<td>S&amp;P 500</td>
<td>3,797.34</td>
<td>+44.59</td>
<td>+1.19%</td>
<td>2.589B</td>
<td>3,741.65</td>
<td>3,810.74</td>
<td>3,491.58</td>
</tr>
<tr>
<td>^DJI</td>
<td>Dow Jones Industrial Average</td>
<td>31,499.62</td>
<td>+417.06</td>
<td>+1.34%</td>
<td>345.036M</td>
<td>31,161.41</td>
<td>31,600.63</td>
<td>28,660.94</td>
</tr>
<tr>
<td>^IXIC</td>
<td>NASDAQ Composite</td>
<td>10,952.61</td>
<td>+92.90</td>
<td>+0.86%</td>
<td>4.063B</td>
<td>10,713.33</td>
<td>10,983.52</td>
<td>10,598.83</td>
</tr>
<tr>
<td>^NYA</td>
<td>NYSE COMPOSITE (DJ)</td>
<td>14,226.11</td>
<td>+82.05</td>
<td>+0.58%</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>^XAX</td>
<td>NYSE AMEX COMPOSITE INDEX</td>
<td>4,295.57</td>
<td>-106.83</td>
<td>-2.43%</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>^BUK100P</td>
<td>Cboe UK 100</td>
<td>701.69</td>
<td>+5.39</td>
<td>+0.77%</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>^RUT</td>
<td>Russell 2000</td>
<td>1,748.40</td>
<td>+6.16</td>
<td>+0.35%</td>
<td>0</td>
<td>1,727.20</td>
<td>1,751.45</td>
<td>1,641.47</td>
</tr>
<tr>
<td>^VIX</td>
<td>Vix</td>
<td>29.85</td>
<td>+0.16</td>
<td>+0.54%</td>
<td>0</td>
<td>29.75</td>
<td>30.95</td>
<td>14.73</td>
</tr>
</tbody>
</table>
import io
import requests
import pandas as pd

response = requests.get('https://finance.yahoo.com/world-indices/')
df = pd.read_html(io.StringIO(response.text))
worldidx = df[0]
worldidx.to_csv('world_indices.csv')
worldidx
# ^GSPC S&P 500
# ^DJI Dow 30
# ^IXIC Nasdaq

!pip install ffn
import ffn
%pylab inline

df = ffn.get('^gspc, ^dji, ^ixic', start='2010-01-01', end='2022-01-01')
print(df.head())
print(df.tail())
print(df.describe())

ax = df.plot(figsize=(12,9))
df = ffn.get('^gspc, ^dji, ^ixic', start='2010-01-01', end='2022-01-01')
ax = df.plot(figsize=(12,9))
!pip install ffn
import ffn
%pylab inline
df = ffn.get('^gspc, ^dji, ^ixic', start='2010-01-01', end='2022-01-01')
print(df.head())
print(df.tail())
print(df.describe())
ax = df.plot(figsize=(12,9))

returns = df.to_returns().dropna()
ax = returns.hist(figsize=(14, 10))
returns.corr().as_format('.2f')
returns.plot_corr_heatmap()
ax = df.plot(figsize=(14,10))

perf = df.calc_stats()
perf.plot(figsize=(14, 10))
print(perf.display())
Normative Theories Revisited

- Revisits the normative theories and analyzes them based on real financial time series data
- Expected Utility and Reality
- Mean-Variance Portfolio Theory (MVPT)
- Capital Asset Pricing Model (CAPM)
- Arbitrage Pricing Theory (APT)

Normalized financial time series data

Simulated portfolio volatilities, returns, and Sharpe ratios

Expected versus realized portfolio volatilities

Expected vs. Realized Portfolio Volatility

Expected versus realized portfolio returns

Expected versus realized portfolio Sharpe ratios

CAPM-predicted versus realized stock returns for a single stock

Average CAPM-predicted versus average realized stock returns for multiple stocks

## Arbitrage Pricing Theory (APT)

### Relevant types of financial data

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market</td>
<td>MSCI World Gross Return Daily USD (PUS = Price Return)</td>
</tr>
<tr>
<td>Size</td>
<td>MSCI World Equal Weight Price Net Index EOD</td>
</tr>
<tr>
<td>Volatility</td>
<td>MSCI World Minimum Volatility Net Return</td>
</tr>
<tr>
<td>Value</td>
<td>MSCI World Value Weighted Gross (NUS for Net)</td>
</tr>
<tr>
<td>Risk</td>
<td>MSCI World Risk Weighted Gross USD EOD</td>
</tr>
<tr>
<td>Growth</td>
<td>MSCI World Quality Net Return USD</td>
</tr>
<tr>
<td>Momentum</td>
<td>MSCI World Momentum Gross Index USD EOD</td>
</tr>
</tbody>
</table>

```python
factors = pd.read_csv('http://hilpisch.com/aiif_eikon_eod_factors.csv',
                      index_col=0, parse_dates=True)
```
APT-predicted versus realized stock returns for a stock

Average APT-predicted versus average realized stock returns for multiple stocks

Normalized factors time series data

APT-predicted returns based on typical factors compared to realized returns

APT-predicted performance and real performance over time (gross)

Debunking Central Assumptions in Finance

• Debunks two of the most commonly found assumptions in financial models and theories
  • Normality of returns
  • Linear relationships

Standard normally distributed random numbers

Distribution with first and second moment of 0.0 and 1.0, respectively

Histogram and PDF for standard normally distributed numbers

Histogram and normal PDF for discrete numbers

Q-Q plot for standard normally distributed numbers

Q-Q plot for discrete numbers

Frequency distribution and normal PDF for S&P 500 log returns

Q-Q for S&P 500 log returns

Expected CAPM return versus beta (including linear regression)

Expected CAPM return versus beta (including linear regression)

Theory-First to Data-Driven Finance

• Finance used to be characterized by normative theories based on simplified mathematical models of the financial markets, relying on assumptions such as normality of returns and linear relationships.

• The almost universal and comprehensive availability of (financial) data has led to a shift in focus from a theory-first approach to data-driven finance.

• Several examples based on real financial data illustrate that many popular financial models and theories cannot survive a confrontation with financial market realities.

• Although elegant, they might be too simplistic to capture the complexities, changing nature, and nonlinearities of financial markets.

The Quant Finance PyData Stack

- PyThalesians
- Zipline
- DX Analytics
- PyAlgoTrade
- QuantLib

- Quantopian
- PyTables
- NetworkX
- scikits-image
- StatsModels
- matplotlib
- pandas

- SciPy
- NumPy
- Python
- SymPy
- PyMC
- Jupyter

Source: http://nbviewer.jupyter.org/format/slides/github/quantopian/pyfolio/blob/master/pyfolio/examples/overview_slides.ipynb#5
Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly

Source: https://www.amazon.com/Artificial-Intelligence-Finance-Python-Based-Guide/dp/1492055433
Yves Hilpisch (2020), *Artificial Intelligence in Finance: A Python-Based Guide*, O’Reilly

https://github.com/yhilpisch/aiif

**Artificial Intelligence in Finance**

**About this Repository**

This repository provides Python code and Jupyter Notebooks accompanying the *Artificial Intelligence in Finance* book published by O’Reilly.
Yves Hilpisch (2020), Artificial Intelligence in Finance: A Python-Based Guide, O’Reilly

https://github.com/yhilpisch/aiif/tree/main/code
Python in Google Colab (Python101)

https://colab.research.google.com/drive/1FEG6DnGvwfUbeo4zJ1zTunjMqf2RkCrT

https://tinyurl.com/aintpupython101
Python in Google Colab (Python101)

Data Driven Finance

Financial Econometrics and Regression

1. import numpy as np
2. def f(x):
3.     return 2 + 1 / 2 * x
4. x = np.arange(-4, 5)
5. array([-4, -3, -2, -1, 0, 1, 2, 3, 4])

6. y = f(x)
7. y
8. array([ 0.00, 0.50, 1.00, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00])

9. print('x', x)
10. print('y', y)
11. beta = np.cov(x, y, ddof=0)[0, 1] / x.var()
12. print('beta', beta)

https://tinyurl.com/aintpuppython101
Normative Theories Revisited

Mean-Variance Portfolio Theory

```python
1 import numpy as np
2 import pandas as pd
3 from pylab import plt, mpl
4 from scipy.optimize import minimize
5 plt.style.use('seaborn')
6 mpl.rcParams['savefig.dpi'] = 300
7 mpl.rcParams['font.family'] = 'serif'
8 np.set_printoptions(precision=5, suppress=True,
9     formatter={'Float': lambda x: f'{x:.3f}'})
10
11 url = 'http://hilpisch.com/aiif_eikon_eod_data.csv'
12 raw = pd.read_csv(url, index_col=0, parse_dates=True).dropna()
13 raw.info()
14
15 symbols = ['AAPL.0', 'MSFT.0', 'INTC.0', 'AMZN.0', 'GLD']
16 rets = np.log(raw[symbols] / raw[symbols].shift(1)).dropna()
17 (raw[symbols].iloc[0].plot(figsize=(10, 6));
18 weights = len(rets.columns) * [1 / len(rets.columns)]
```

https://tinyurl.com/aintpupypton101
Python in Google Colab (Python101)

https://tinyurl.com/aintpupython101
Python in Google Colab (Python101)

https://tinyurl.com/aintpupython101
Python in Google Colab (Python101)

https://tinyurl.com/aintpupython101
Python in Google Colab (Python101)

https://tinyurl.com/aintpupython101
Summary

• Data-Driven Finance
• Scientific Method
• Financial Econometrics and Regression
• Data Availability
• Normative Theories Revisited
• Debunking Central Assumptions in Finance

References


• Min-Yuh Day (2022), Python 101, [https://tinyurl.com/aintpupython101](https://tinyurl.com/aintpupython101)