Data Driven Methods in Finance:

Fundamental Factor Models

Fall 2023: IEOR 4576

Naftali Cohen

Introduction

The central idea of modern financial economics:

The average return of a stock is the payoff for taking   risk.

Introduction

The central idea of modern financial economics:

The average return of a stock is the payoff for taking *non-diversifiable* risk.

Introduction

The return on stock i can be written as $$r_i = \alpha + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i$$

The average stock return is $$E(r_i) = \underbrace{E(\alpha)}_{\alpha} + \beta_{i1} E(f_1) + \cdots + \beta_{iK} E(f_K)$$

This can be written as $$r_i = \alpha + \boldsymbol{\beta}_i' \mathbf{f} + \epsilon_i$$

And the average stock return becomes $$E(r_i) = \alpha + \boldsymbol{\beta}_i' E(\mathbf{f})$$

Introduction

Average stock return = factor exposure x factor premium

$$E(r_i) = \alpha + \boldsymbol{\beta}_i' E(\mathbf{f})$$

- Factor exposure (beta) represents the exposure of a stock to some kind of risk.
- The factor premium (f) quantifies the payoff to an investor who takes on that risk by buying the stock.
- The average stock return is the payoff for taking risk—**but what is this risk exactly?**

Introduction

The risk of a stock has two components, diversifiable risk and nondiversifiable risk.

Total risk = nondiversifiable risk + diversifiable risk

Since investors can eliminate diversifiable risk from their portfolios through diversification, the market only rewards exposure to nondiversifiable risk.

The average stock return is the payoff for taking nondiversifiable risk.

Introduction

nondiversifiable risk = factor exposure² x factor risk

$$
\begin{aligned}
V(r_i) &= V(\alpha + \beta_{i1}f_1 + \cdots + \beta_{iK}f_K) + V(\epsilon_i) \\
&= V(\beta_{i1}f_1 + \cdots + \beta_{iK}f_K) + V(\epsilon_i)
\end{aligned}
$$

$$
V(r_i) = \boldsymbol{\beta}_i' \, V(\mathbf{f})\boldsymbol{\beta}_i + V(\epsilon_i)
$$

Creating a Fundamental model: preliminary work

Preliminary work:

1. Choose the factors for the model.
2. Determine the treatment of the risk-free rate.
3. Define the investment universe.
4. Decide on the time interval and time period of the data.

Creating a Fundamental model: preliminary work

Choose the factors for the model.

- Factors represent risk.
- Fundamental factors are observable characteristics of the stock itself. For example,
  - valuation factor: earnings-to-price ratio, book-to-price ratio
  - size factor: log of market capitalization
  - analyst factors: analyst rating changes, earnings revisions
  - technical factors: 12-month momentum, trading volume

Creating a Fundamental model: preliminary work

Determine the treatment of the risk-free rate.

- How much of a stock's return comes from the stock itself, and how much could have been earned on **any** investment?
- The investor earns a portion of the average stock return **for free—not as a reward for taking on risk.**
- The reward for taking on risk by buying a stock is therefore the average stock return in excess of a risk-free rate.

$$r_{it}^{*} = r_{it} - r_{ft}$$

Creating a Fundamental model: preliminary work

Define the investment universe.

- Portfolio managers must decide on the number of stocks to include in their model, considering possible external limitations or personal screening strategies.
- Modern computing allows handling vast numbers of stocks effortlessly, so technological constraints aren't a concern.
- While a larger investment universe increases the pool of potential high-return stocks, calculating correlations becomes less precise with more stocks, potentially reducing model reliability.

Creating a Fundamental model: preliminary work

Decide on the time interval and time period of the data.

~60 data points

- **The time interval between data points should ideally align with the investment horizon** (e.g., monthly rebalancing should use monthly intervals), but considerations like stability of relationships and estimation precision might dictate otherwise.
- While a shorter time interval is typically preferred to avoid issues with relationship **stability**, the overall time period (or sample period) should ensure enough data points for precision, but not so long that relationships change significantly or data collection becomes cumbersome.
- Financial economists often favor monthly intervals with a total time period of three to five years, adjusting the number of intervals based on the chosen frequency.

Creating a Fundamental model: main steps for modeling

Main steps for modeling total stock return (for *non-benchmark* managers):

1. Stock returns are estimated with the following equation: $r_i = \alpha + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i.$
2. Collect stock-return and factor exposure data for the time period at each time interval.
3. Estimate the factor premium from a panel regression of the stock return on the factor exposure using either the OLS, MAD, or the GLS method.
4. Check the robustness of the factor premium estimation by splitting the data into subsets and comparing the estimates for each subset. If the estimates are similar across subsets, the estimation is robust.
5. If the estimation is not robust, try a different estimation method.
6. Calculate each stock’s average return as the product of the factor premium and the vector containing the stock’s factor exposures.
7. Decompose the risk of the stock return into its diversifiable and nondiversifiable components.
8. Calculate the correlation between the returns of the stocks in the investment universe.

Creating a Fundamental model: main steps for modeling

Benchmark and α:

Many portfolio managers have a specific benchmark against which they measure the performance of their portfolios. If the portfolio manager aims to (1) outperform a benchmark while (2) minimizing the portfolio's tracking error, then the benchmark must play a role in the model.

Approaches:

- use the model to predict only the residual return rather than the entire stock return.
- add the benchmark to the model as one of the factors.
- controlling for the tracking error in construction of the portfolio.

Creating a Fundamental model: main steps for modeling

Main steps for modeling total stock return (for *benchmark* managers):

1. Stock returns are estimated with the following equation: $r_i = \tilde{\alpha}_i + \tilde{\beta}_i r_B + \tilde{\epsilon}_i.$
2. Collect stock return and benchmark return data for the given time period at the given time intervals.
3. After estimating $\tilde{\alpha}$ and $\tilde{\beta}_i$ calculate the residual return as $\tilde{r}_i \equiv \tilde{\alpha}_i + \tilde{\epsilon}_i.$
4. Estimate residual return with the following equation: $\tilde{r}_i = \alpha + \beta_{i1} f_1 + \cdots + \beta_{iK} f_K + \epsilon_i.$
5. Follow steps 2 through 8 of the main steps for modeling total stock return, substituting the residual return for the total stock return.

Creating a Fundamental model: main steps for modeling

Factor Exposure:

- The factor exposure quantifies the exposure of a stock to risk.
- In the fundamental factor model, the factor exposure is the value of some observable (financial statements, price-volume chart, or easily calculable) characteristic of the stock.
- Factor exposures change over time, and we need to be sure to assign the correct values to the correct dates.
- Generally, it is reasonable to allow two to three months' lag between the end of a fiscal period and the reporting of the variable.

Creating a Fundamental model: main steps for modeling

Factor Exposure:

Factor Exposure of Selected Stocks

Ticker	E/P	B/P	D/E	LOGSIZE	M12M
AAPL	0.0289	0.0357	3.3904	14.5206	0.7973
MSFT	0.0274	0.0731	1.5469	14.2970	0.4291
AMZN	0.0083	0.0464	2.5036	14.2790	0.7592
GOOGL	0.0285	0.1871	0.3433	13.9183	0.3493
FB	0.0353	0.1659	0.2648	13.4088	0.3736
BRK.B	0.0414	0.7329	0.9840	13.1935	0.0391
WMT	0.0413	0.1739	1.9235	12.9784	0.3056
JNJ	0.0399	0.1654	1.5148	12.8502	0.0812
JPM	0.0700	0.6523	11.1494	12.7920	−0.0727
V	0.0330	0.0849	1.1924	12.7827	0.1471

Creating a Fundamental model: main steps for modeling

Factor Exposure (PIT guidelines):

- The factor exposure for a month is derived from the latest data at the start of that month.
- For quarterly data, a reporting gap of three months is assumed.
- For annual data, if the fiscal year ends in December, then data from the last fiscal year available in December 2019 is used.

Data Frequency	If the Fiscal Quarter Ends in:	If the Fiscal Year Ends in:	Use the Data for:
Daily data			End of Nov 2020
Monthly data			Nov 2020
Quarterly data	Mar, Jun, Sep, Dec		Apr 2020–Jun 2020
	Feb, Mar, Aug, Nov		Jun 2020–Aug 2020
	Jan, Apr, Jul, Oct		May 2020–Jul 2020
Annual data		Dec	Jan 2019–Dec 2019
		Nov	Dec 2018–Nov 2019
		Oct	Nov 2018–Oct 2019
		Sept	Oct 2018–Sep 2019
		Aug	Sep 2019–Aug 2020
		⋮	⋮

Creating a Fundamental model: main steps for modeling

Factor Premium:

- Factor premium is the payoff for each unit of factor exposure, or exposure to risk, that the stock possesses.
- In the fundamental factor model, the factor premium is estimated from the pooled cross-sectional regression (i.e., panel regression) of the stock return on the factor exposure.
- Estimation of the factor premium with a regression is possible because the premium likely remains constant over several years and across stocks.

$$r_{it} = \alpha + \beta_{it}'\mathbf{f} + \epsilon_{it}$$

Creating a Fundamental model: main steps for modeling

Factor Premium:

- OLS estimation of the factor premium: $\hat{\mathbf{f}} = \left[\sum_{t=1}^{T} \sum_{i=1}^{N} (\beta_{it} - \bar{\beta})(\beta_{it} - \bar{\beta})' \right]^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} (\beta_{it} - \bar{\beta})(r_{it} - \bar{r})$
- With factor premium variance: $\hat{V}(\hat{\mathbf{f}}) = \hat{\sigma}^2 \left[\sum_{t=1}^{T} \sum_{i=1}^{N} (\beta_{it} - \bar{\beta})(\beta_{it} - \bar{\beta})' \right]^{-1}$

E/P	B/P	D/E	LOGSIZE	M12M
−3.4248 (0.4051)	0.2229 (0.1652)	−0.0005 (0.0008)	0.0470 (0.0549)	−2.4032 (0.2022)

Note: Estimated for the period from January 2016 to December 2020. Standard errors are inside parentheses. Factors are as in Table 6.2.

short-term vs.

long-term value

failure?

Creating a Fundamental model: robustness check

- Models aim to approximate reality, with any discrepancies termed as specification errors. Robustness checks assess the stability of model estimates when small estimation details vary.
- For a robust estimation, factor premium estimates shouldn't vary significantly across different data subsets.
- Robustness can be checked by dividing the data over time, like comparing January 2016-June 2018 with July 2018-December 2020, or across sectors.

The Factor Premium for Subperiods

Period	E/P	B/P	D/E	LOGSIZE	M12M
Jan 2016–Jun 2018	−2.5101	0.3373	−0.0005	−0.0319	−0.9915
	(0.4041)	(0.1960)	(0.0007)	(0.0628)	(0.2225)
Jul 2018–Dec 2020	−5.1183	−0.2026	−0.0044	0.1601	−4.1142
	(0.7909)	(0.2657)	(0.0047)	(0.0902)	(0.3497)

Creating a Fundamental model: robustness check

Sector	E/P	B/P	D/E	LOGSIZE	M12M
10	−4.8850	2.0557	0.0455	0.1761	−5.6620
	(0.9017)	(0.7182)	(0.1483)	(0.3644)	(1.1626)
15	−10.8167	−0.0625	−0.0157	−0.4396	−3.1329
	(2.0278)	(0.8931)	(0.0828)	(0.3454)	(0.9029)
20	−1.6735	0.7879	−0.0038	−0.0841	−3.0761
	(2.6884)	(0.8404)	(0.0046)	(0.1517)	(0.5620)
25	−4.2285	1.5400	−0.0003	0.3712	−2.4307
	(2.1720)	(0.6386)	(0.0009)	(0.1708)	(0.5544)
30	5.5367	−1.2432	−0.0023	−0.0827	−3.4636
	(2.2242)	(0.6087)	(0.0038)	(0.1641)	(0.7859)
35	5.0186	−1.6256	0.0202	−0.1480	−1.7809
	(3.1616)	(0.6732)	(0.0332)	(0.1437)	(0.4803)
40	−3.2311	0.4321	0.0122	−0.1228	−3.9668
	(3.3614)	(0.3431)	(0.0234)	(0.1343)	(0.6167)
45	−4.5227	−0.7912	0.0023	0.1103	−1.8753
	(2.2509)	(0.6619)	(0.0062)	(0.1250)	(0.4332)
50	−0.4454	0.5284	−0.0977	0.6444	−6.7083
	(3.3899)	(1.1857)	(0.1504)	(0.2827)	(1.6776)
55	−0.6537	0.3741	0.3129	−0.2492	−2.6367
	(0.9968)	(0.9778)	(0.1422)	(0.2326)	(0.8356)

Note: Estimated for the period from January 2016 to December 2020. Standard errors are inside parentheses. Factors are as in Table 6.2. Sector names are based on the Global Industry Classification Standard (GICS), i.e., Energy (10), Materials (15), Industrials (20), Consumer Discretionary (25), Consumer Staples (30), Health Care (35), Financials (40), Information Technology (45), Telecommunication Services (50), and Utilities (55).

Creating a Fundamental model: Outliers and LAD

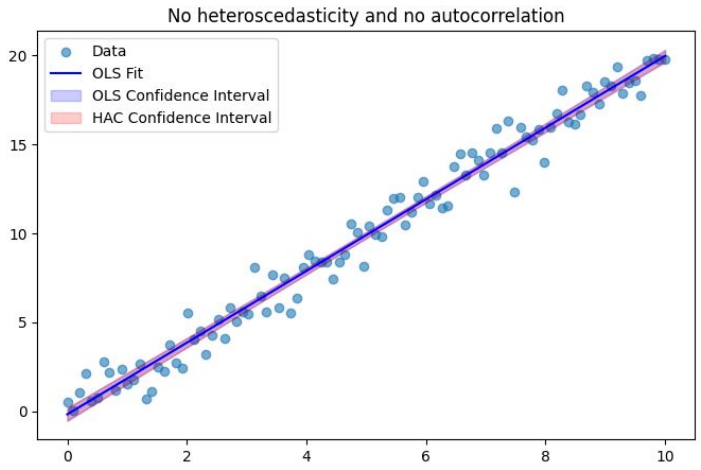
- The Ordinary Least Squares (OLS) method is sensitive to outliers, potentially leading to unstable estimations.
- The Least Absolute Deviation (LAD) or Minimum Absolute Deviation (MAD) estimation reduces sensitivity to outliers by **minimizing the absolute value of residuals**, unlike OLS which **minimizes squared residuals**.
- In a provided example, using the MAD approach changed the earnings-to-price ratio (E/P) from significantly negative to positive, suggesting outliers heavily influenced the OLS estimate.

The MAD Factor Premium

E/P	B/P	D/E	LOGSIZE	M12M
0.7784 (0.3639)	−0.2912 (0.1484)	−0.0004 (0.0007)	−0.0085 (0.0493)	−2.6169 (0.1817)

Note: Estimated for the period from January 2016 to December 2020. Standard errors are inside parentheses. Factors are as in Table 6.2.

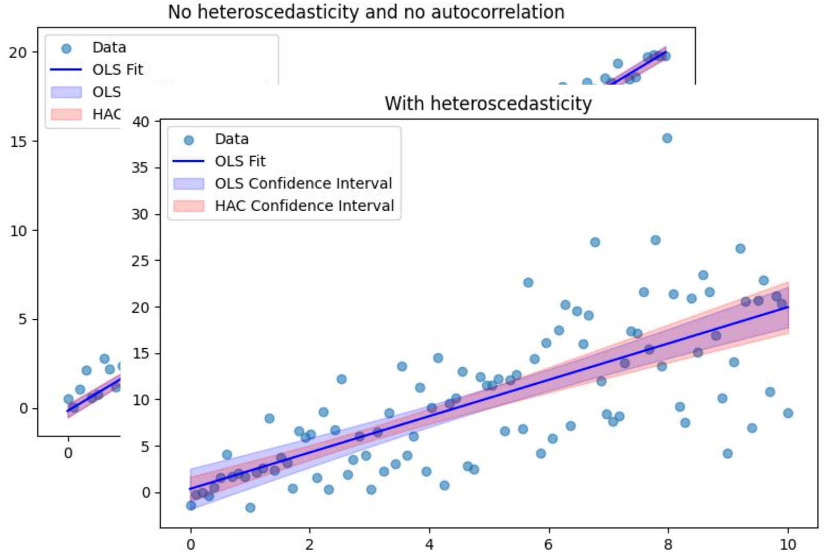
Creating a Fundamental model: HAC



x = np.linspace(0, 10, 100)

*y = 2*x + np.random.normal(0, 1, len(x))*

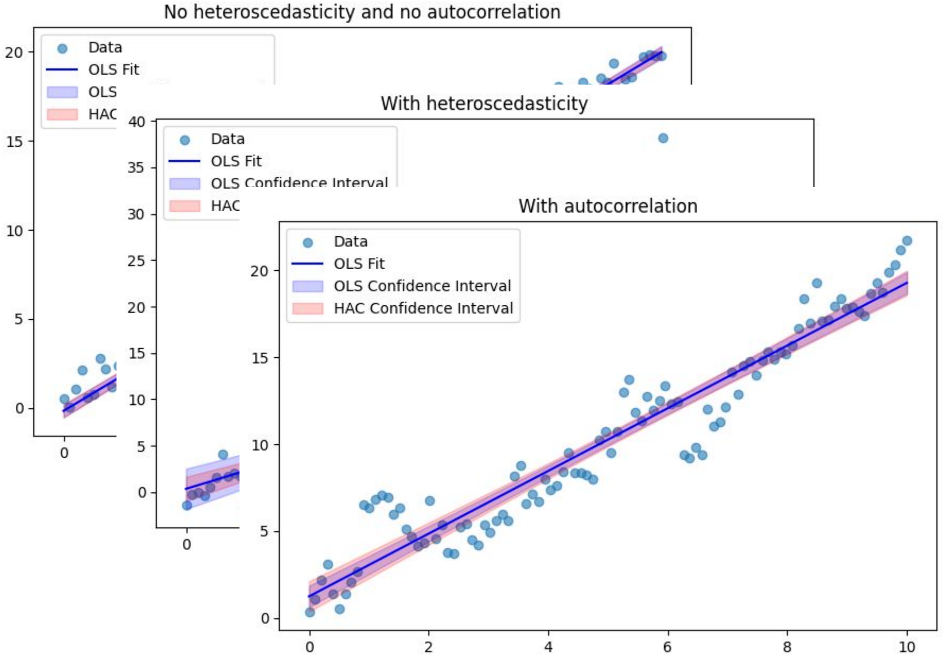
Creating a Fundamental model: HAC




```
x = np.linspace(0, 10, 100)

error_scale = np.linspace(1, 10, 100)
y = 2*x + np.random.normal(0, error_scale, len(x))
```

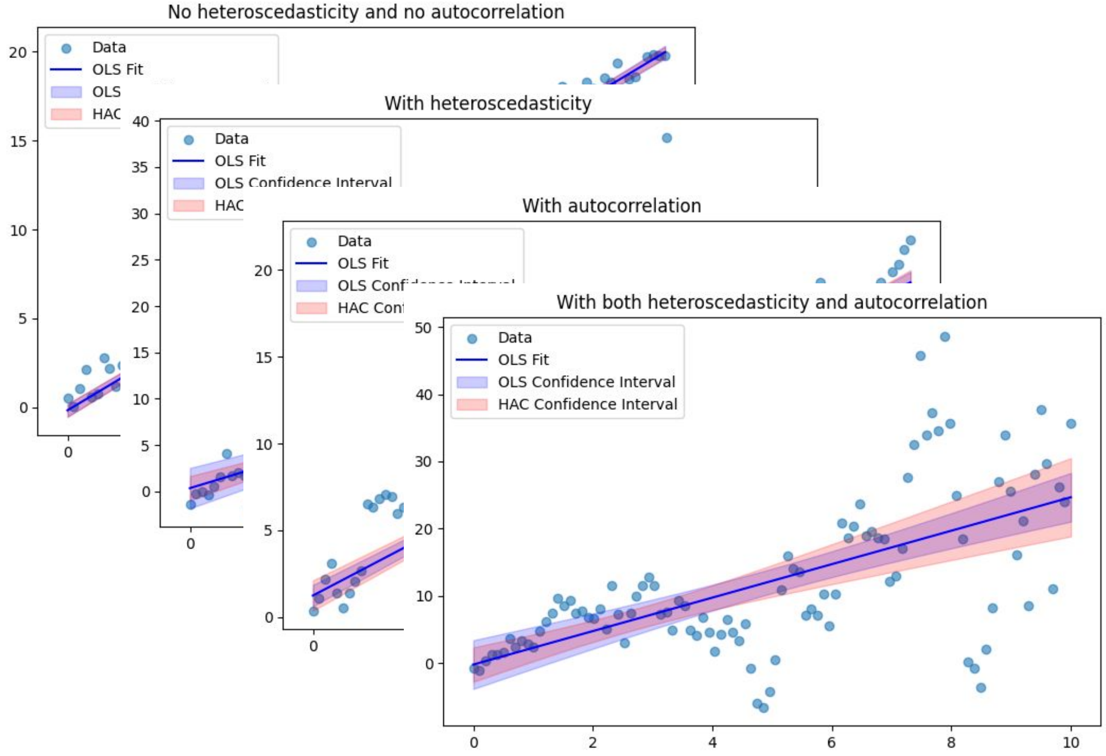
Creating a Fundamental model: HAC




```
x = np.linspace(0, 10, 100)

errors = [np.random.normal(0, 1)]
for i in range(1, len(x)):
    errors.append(0.8*errors[i-1] + np.random.normal(0, 1))
y = 2*x + np.array(errors)
```

Creating a Fundamental model: HAC



$$Var(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1}$$

$$Var(\hat{\beta}_{HAC}) = (X'X)^{-1} X' \Omega X (X'X)^{-1}$$

$$\hat{\Omega} = \sum_{i=-\infty}^{\infty} \omega_i \hat{\epsilon}_t \hat{\epsilon}'_{t-i}$$

$$SE_{HAC}(\hat{\beta}) > SE_{OLS}(\hat{\beta})$$

Creating a Fundamental model: HAC

- OLS (Ordinary Least Squares) is a standard empirical tool for data analysis, but its output can sometimes be unreliable due to **certain assumptions**.
- To improve the reliability of OLS, adjustments can be made to the standard errors by relaxing two assumptions: (1) **homoscedastic errors** (equal error variances across firms) and (2) **no autocorrelation** (independence of error terms across time periods).
- The adjusted standard errors, which account for heteroscedasticity and autocorrelation, are termed HAC (heteroscedasticity- and autocorrelation-consistent) standard errors. These errors are based on a "sandwich formula" that factors in variations of both factor exposure and variance.

Creating a Fundamental model: HAC

- A comparison of HAC standard errors with unadjusted ones indicates that certain factor premiums, like the earnings-to-price and book-to-price ratios, may be less precisely estimated than initially thought.

The HAC Standard Errors along with the OLS Estimates

E/P	B/P	D/E	LOGSIZE	M12M
−3.4248	0.2229	−0.0005	0.0470	−2.4032
(1.1315)	(0.2268)	(0.0003)	(0.0612)	(0.3552)

Note: Estimated for the period from January 2016 to December 2020. Standard errors are inside parentheses. Factors are as in Table 6.2.

Creating a Fundamental model: Risk decomposition

Nondiversifiable risk arises from the randomness of the factor premium.

$$\frac{1}{T}\sum_{t=1}^{T}\left[\hat{\alpha}_t + \beta_{it}'\hat{\mathbf{f}}_t - \frac{1}{T}\sum_{t'=1}^{T}\left(\hat{\alpha}_{t'} + \hat{\beta}_{it'}'\hat{\mathbf{f}}_{t'}\right)\right]^2$$

The **diversifiable risk** represents the part of the variation in the stock return that the variation in the model's factors cannot explain.

$$\hat{\sigma}_i^2 = \frac{1}{T}\sum_{t=1}^{T}\left(r_{it} - \hat{\alpha}_t - \beta_{it}'\hat{\mathbf{f}}_t\right)^2$$

The **total risk** of stock i is simply the sum of the two risk components

Creating a Fundamental model: Risk decomposition

The Risk Decomposition for Selected Stocks

Ticker	Total Risk		Nondiversifiable Risk		Diversifiable Risk	
	V	SD	*V*	SD	*V*	SD
AAPL	824.50	28.71	170.47	13.06	654.03	25.57
MSFT	364.38	19.09	149.52	12.23	214.85	14.66
AMZN	667.40	25.83	183.24	13.54	484.17	22.00
GOOGL	438.11	20.93	167.99	12.96	270.12	16.44
FB	682.49	26.12	179.62	13.40	502.86	22.42
BRK.B	368.75	19.20	254.25	15.95	114.51	10.70
JNJ	336.23	18.34	178.93	13.38	157.30	12.54
WMT	450.71	21.23	159.11	12.61	291.60	17.08
JPM	456.13	21.36	272.33	16.50	183.80	13.56
V	275.62	16.60	158.06	12.57	117.55	10.84

Note: Based on the period from January 2016 to December 2020. *V* refers to the variance, and SD refers to the standard deviation. Both variance and standard deviation are annualized and expressed in percentage. Company names for ticker symbols are reported in Table 6.2.

Creating a Fundamental model: Risk decomposition

While finding an optimal portfolio, correlations between stock returns are essential. These correlations involve both nondiversifiable and diversifiable components.

However, in practice, only the correlation between the nondiversifiable components is typically estimated, often because estimating diversifiable components is complex and deemed less impactful on overall stock return.

$$C(r_{it}, r_{jt}) = C\left(\hat{\beta}'_{it}\hat{\mathbf{f}}_t, \hat{\beta}'_{jt}\hat{\mathbf{f}}_t\right) + C(\hat{\epsilon}_{it}, \hat{\epsilon}_{jt})$$

Disclaimer

This course is for educational purposes only and does not offer investment advice or pre-packaged trading algorithms. The views expressed herein are not representative of any affiliated organizations or agencies. The main objective is to explore the specific challenges that arise when applying Data Science and Machine Learning techniques to financial data. Such challenges include, but are not limited to, issues like short historical data, non-stationarity, regime changes, and low signal-to-noise ratios, all of which contribute to the difficulty in achieving consistently robust results. The topics covered aim to provide a framework for making more informed investment decisions through a systematic and scientifically-grounded approach.