Big Data and Machine Learning in Asset Management
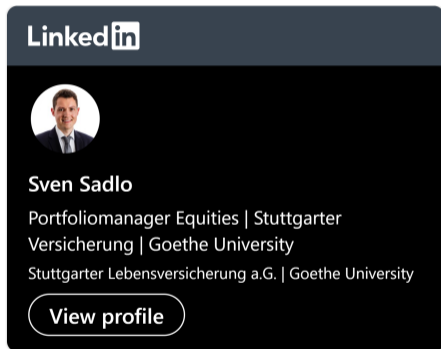
Sven Sadlo

Goethe-University, Frankfurt am Main

December 13, 2021

The views, thoughts, and opinions expressed in this presentation are solely my own, and not necessarily those of any other organization that is mentioned.
For educational and informational purposes only. No investment advice.

Introduction Sven Sadlo

Linkedin

Sven Sadlo

Portfoliomanager Equities | Stuttgarter
Versicherung | Goethe University

Stuttgarter Lebensversicherung a.G. | Goethe University

View profile

Education

2020 Goethe-University, Frankfurt am Main
Money and Finance M. Sc.

2018 University of Hohenheim, Stuttgart
Business Admin. & Economics B. Sc.

Experience

2021 AgnosticInvesting.com
Website, Blog, and Wikifolios

2020 Stuttgarter Lebensversicherung, Stuttgart
Junior Portfoliomanager – Equities

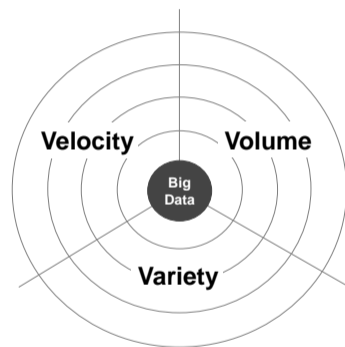# What is Big Data?[1]

No clear definition ...

► Buzzword for data, technology and more.

► Framework: 3 Vs by Doug Laney (2001).

► Data beyond "standard" methodology.

Big Data for this talk:

► Data that requires new methodology.

► "Variety" most interesting for finance.

Own illustration of Sagiroglu and Sinanc (2013).

[1] Sources: Goldstein et al. (2021), Martin and Nagel (2019).

What is Machine Learning?[2]

A very broad development ...

- ► Progress from data and computing power.

- ► Today: collection of algorithms and models.

- ► Models overcome previous limitations.

Machine Learning for this talk:
Predictive models that **"learn"** over time and can handle **many inputs** and **non-linearities**.

*Moreover, a large part of what is branded AI (or ML) in finance is **not new** but has existed in the form of **statistical or econometric modeling** for a long time.*

Bartram et al. (2020, p.4)

---
[2] Sources: Rasekhschaffe and Jones (2019), Bartram et al. (2020, 2021), Gu et al. (2020), Israel et al. (2020), Leung et al. (2021), Nagel (2021).

Machine Learning vs. Econometrics

Traditional Econometrics

► Goal: identify statistically significant effects!

► Driven by theory and hypotheses.

► Forecasting is secondary.

$$\Downarrow$$

Research

Machine Learning

► Goal: best possible forecast out-of-sample!

► Driven by data and unknown relations.

► Marginal effects are secondary.

$$\Downarrow$$

Practice

What is Asset Management? A process[3] ...

1. Return Prediction and Portfolio Construction
▶ Which securities and how to combine them?
▶ Focus: finding and maintaining profitable signals!

2. Implementation, Trading, and Execution
▶ How to buy and sell the securities efficiently?
▶ Challenge: fast execution while controlling costs!

3. Risk Management and Portfolio Monitoring
▶ How to prevent the worst case?
▶ Challenge: preserve the strategy while controlling risks!

---

[3] Inspired by Bartram et al. (2020, 2021).

Can Machines "Learn" Finance?[4]

Problem #1: Small Data
► There is (often) not enough data available for machine learning.
► And to get more you have to wait …

Problem #2: Low Signal-to-Noise Ratios
► Competition in financial markets (often) eliminates predictability.
► And there is a lot of randomness anyway …

Problem #3: Evolving Environment
► There are (almost) no fixed rules or relations in financial markets.
► Even worse: things that worked stop working because they worked …

---

[4] Source: Israel et al. (2020).

Realistic expectations are important ...

	Problem #1	Problem #2	Problem #3
	Small Data	Signal-to-Noise Ratio	Evolving Environment
Return Prediction and Portfolio Construction	🔴	🔴	🔴
Implementation, Trading, and Execution	🟢	🟡	🔴
Risk Management and Portfolio Monitoring	🔴	🟡	🔴

Hardest but also most important → Focus of remaining talk!

The heart of active portfolio management ...

Inspired by Pedersen (2015, Figure 3.1).

How to evaluate investment processes?

		Problems with Machine Learning in Finance		
		Problem #1 Small Data	**Problem #2** Signal-to-Noise Ratio	**Problem #3** Evolving Environment
How to Beat the Market?	**Machine Learning** Novel Analysis of Existing Information			
	Alternative Data Production of New Information			

Real-World Strategies and Funds – Overview

Alternative Data[5]

► ACATIS AI Buzz US Equities

► LI Data Intelligence Germany

► ART AI Europe

---

[5]Disclaimer: for educational and informational purposes only. No investment advice.

ACATIS AI Buzz US Equities

Sentiment Analysis – USA

▶ Fund tracks the BUZZ Index created by Periscope Capital.

▶ Idea: investor sentiment is important!

▶ Data: online chatter from social media, news, etc. (>15m posts per month).

▶ Sophisticated NLP classifies content: bullish, neutral, bearish.

▶ Invests in 75 stocks with most bullish sentiment. Rebalanced monthly.

Source: fund website and company representatives.
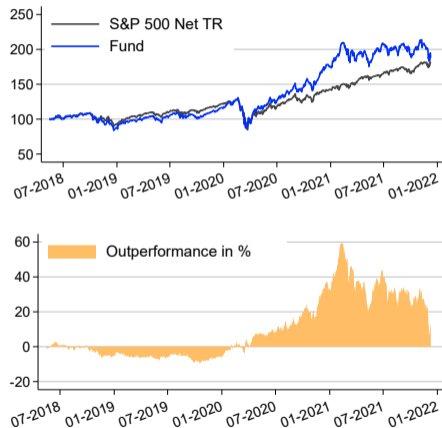
Data from Bloomberg. Returns in USD from 2018-05-15 to 2021-12-09.

Lehner Investments – LI Data Intelligence Germany

Sentiment Analysis – Germany

► Similar strategy but for German stocks.

► Company crawls 2m messages per day from *"hundreds of sources"*.

► NLP classifies content in 3 languages (German, English, Chinese).

► Invests in 30 stocks with most bullish signal. Rebalanced every 3 weeks.

Source: fund website and company representatives.
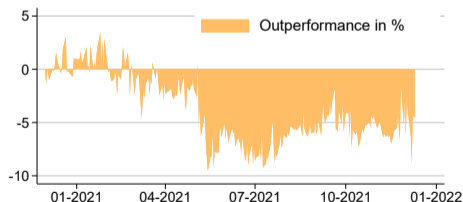
Data from Bloomberg. Returns in EUR from 2020-11-30 to 2021-12-10.

Evaluating the investment processes ...

		Problems with Machine Learning in Finance		
		Problem #1 Small Data	**Problem #2** Signal-to-Noise Ratio	**Problem #3** Evolving Environment
How to Beat the Market?	**Machine Learning** Novel Analysis of Existing Information	Not used.	Not used.	Not used.
	Alternative Data Production of New Information	**No Problem!** Millions of posts, news, articles, etc. available.	**No Problem!** Text classification well established in ML.	**Problem!** Portfolio construction is static but role of sentiment changes.

QI Investment – ART AI Europe

Information from Alternative Data

► Market neutral equity fund: stock selection and index hedge.

► Benchmark: constant positive return.

► *"Key driver of the strategy is data from new sources."*

► Examples for Alternative Data:

　► Satellite images
　► Job postings
　► App downloads and web traffic

► Not much information available.

Data from Bloomberg. Returns in EUR from 2019-07-23 to 2021-12-09.

Source: fund website and company representatives.

Real-World Strategies and Funds – Overview

Machine Learning[6]

- ► ACATIS AI Global Equities

- ► HQAM European Equities

- ► Z22 Smart Mirror

- ► Castle Ridge Asset Management

---

[6]Disclaimer: for educational and informational purposes only. No investment advice.

ACATIS AI Global Equities

Value Investing by the Machine

► Idea: ML for fundamental analysis!

► Data: conference call transcripts, fundamentals, and sector indicators.

► First attempt: automated process with lots of features and few constraints.

► Refinement: ML only used to identify *winners* within industries.

► Portfolio of 50 stocks that is sector- and country-neutral to the benchmark.

► Rebalanced only every 6 months.

Source: fund website and company representatives.
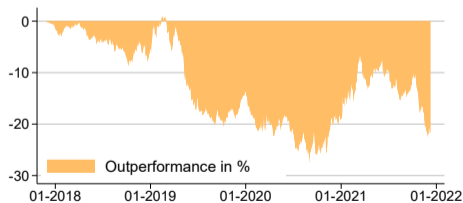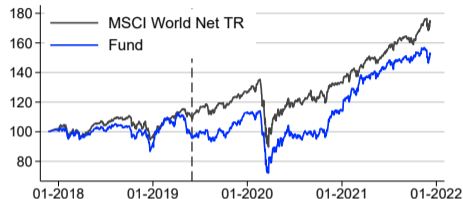
Data from Bloomberg. Returns in EUR from 2017-11-23 to 2021-12-09.

ACATIS AI Global Equities (cont.)

		Problems with Machine Learning in Finance		
		Problem #1 Small Data	**Problem #2** Signal-to-Noise Ratio	**Problem #3** Evolving Environment
How to Beat the Market?	**Machine Learning** Novel Analysis of Existing Information	**Problem!** Fundamentals are low-frequency and small data.	**Problem!** Returns and fundamentals are hard to predict.	**Problem!** Role of fundamentals changes and portfolio only rebalanced semi-annually.
	Alternative Data Production of New Information	**Partly Problematic!** Sample size of call transcripts only in the thousands.	**No Problem!** Text classification well established in ML.	**Partly Problematic!** Feedback effects since firms manage sentiment.

HQ Asset Management – HQAM European Equities

Improving the MSCI Europe

- ► Goal: constant outperformance with limited active risk!

- ► Data: all kind of factors discussed in the literature. No alternative data.

- ► Advanced multi-factor strategy with ML.

- ► ML used for stock selection and portfolio optimization. No timing.

- ► Diversified portfolio of usually >200 stocks. Dynamically rebalanced.
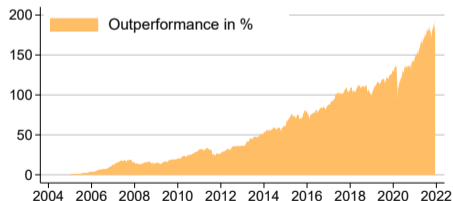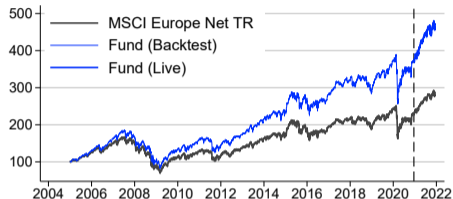
Source: fund website and company representatives.

Data from Bloomberg. Returns in EUR from 2005-01-03 to 2021-12-10.

HQ Asset Management – HQAM European Equities (cont.)

		Problems with Machine Learning in Finance		
		Problem #1 Small Data	**Problem #2** Signal-to-Noise Ratio	**Problem #3** Evolving Environment
How to Beat the Market?	**Machine Learning** Novel Analysis of Existing Information	**Problem!** Most factors are low-frequency and small data.	**Problem!** Returns and factors are hard to predict.	**Problem!** Importance of factors not stable over time.
		Low active risk and narrow focus on stock selection mitigate problems!		
	Alternative Data Production of New Information	Not used.	Not used.	Not used.

Z22 Technologies – Z22 Smart Mirror

Protect the Downside

► Goal: capture equity risk premium while avoiding drawdowns!

► Stock selection based on 13-F filings (50-150 stocks).

► Market timing via volatility forecast and long-volatility ETF.

► Both elements use machine learning.

► Only investable via certificate.

Source: fund website and company representatives.
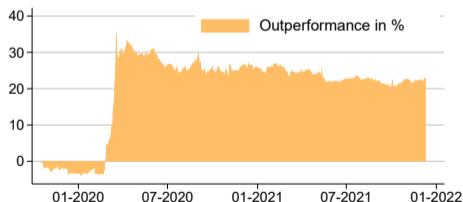
Data from Bloomberg. Returns in USD from 2019-10-15 to 2021-12-10.

Z22 Technologies – Z22 Smart Mirror (cont.)

		Problems with Machine Learning in Finance		
		Problem #1 Small Data	**Problem #2** Signal-to-Noise Ratio	**Problem #3** Evolving Environment
How to Beat the Market?	**Machine Learning** Novel Analysis of Existing Information	**Problem!** Volatility forecasting and stock selection is small data problem.	**Improved!** Research concludes that volatility is easier to predict than returns.	**Problem!** Drivers and behavior of volatility not stable over time.
	Alternative Data Production of New Information	**Partly Problematic!** Large number of holdings in 13-F filings but still driven by return data.	**Partly Problematic!** Returns are hard to predict. Maybe easier to identify good managers.	**Problem!** No stable rules for return prediction and manager classification.

Castle Ridge Asset Management

Meet W.A.L.L.A.C.E.

► Market neutral hedge fund for liquid US stock market (live since 10/2019).

► Proprietary AI platform ("Wallace") focused on evolutionary algorithms.

► Wallace consumes 42 features and achieves accuracy of 72%.

► Wallace found to anticipate idiosyncratic events (e.g. takeovers).

► Dollar- and beta-neutral portfolio of about 100 stocks. Rebalanced weekly.

Source: fund website and company representatives.

Growth of Initial 1,000

Returns (%)　WALLACE MN　HFR EMN　S&P 500 TR
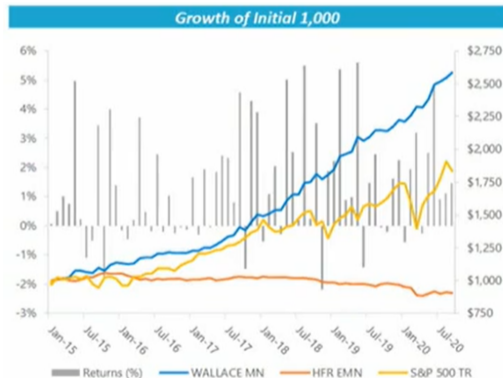
Castle Ridge Asset Management (cont.)

		Problems with Machine Learning in Finance		
		Problem #1 Small Data	**Problem #2** Signal-to-Noise Ratio	**Problem #3** Evolving Environment
How to Beat the Market?	**Machine Learning** Novel Analysis of Existing Information	**Problem!** Return prediction and risk forecasting is small data problem.	**Improved!** Algorithm identifies idiosyncratic anomalies relative to peers.	**Improved!** Evolutionary, adaptive element is key innovation of Wallace AI platform.
	Alternative Data Production of New Information	Not used.	Not used.	Not used.

Real-World Strategies and Funds – Overview

More Systematic Overview[7]

- ► NextGen AI Multi-Manager Index

- ► PLEXUS AI Outperformance Index

- ► Eurekahedge AI Hedge Fund Index

---

[7]Disclaimer: for educational and informational purposes only. No investment advice.

NextGen AI Multi-Manager Index[8]

► Index is *"comprised of AI powered funds whose investment process is underpinned and powered by artificial intelligence [...]"*.

► Equally weighted index of single-manager AI funds (constituents are proprietary). Investable product for the index is planned!


■2nd NextGen Artificial Intelligence Multi-Manager Index

---
[8] Source: NextGen Alpha and Solactive. Index in EUR from 2018-09-28 to 2021-11-30.

PLEXUS Investments AI Outperformance Index[9]

► *"[...] funds that employ AI or ML in their investment process."*

► Index of equally weighted outperformance.

► AI-Outperformer Ratio: fraction of funds that beat their benchmark.

Outperformance	1 Month	3 Months	6 Months	12 Months
Mean	-0.35 %	-2.75 %	-5.09 %	-5.79 %
Top	2.95 %	2.54 %	1.54 %	18.20 %
75 % Quantil	0.62 %	-0.14 %	-1.08 %	2.98 %
Median	-0.41 %	-1.87 %	-2.72 %	-5.27 %
25 % Quantil	-1.11 %	-5.36 %	-8.82 %	-18.43 %
Flop	-4.02 %	-10.55 %	-16.14 %	-27.51 %

Most AI funds behind their benchmark. Nothing new for active managers ...

AI-Outperformer Ratio (1mo)

AI-Outperformer Ratio (3mo)

AI-Outperformance Index
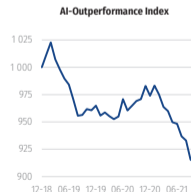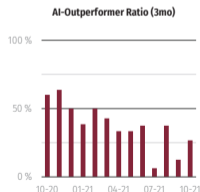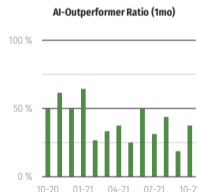
[9]Source: PLEXUS Investments – Special Topic AI as of 10/2021.

Eurekahedge AI Hedge Fund Index

► Equally weighted index of 14 funds.

► Constituents are proprietary.

► Hedge funds that *"utilize artificial intelligence and machine learning theory in their trading processes."*

► Benchmark: broad hedge fund index with currently 2,502 constituents.

► Some remarks on hedge fund indices:

  ► Voluntary and selective reporting.
  ► Non-investable funds included.
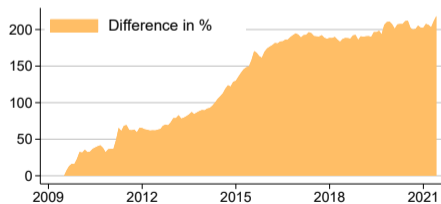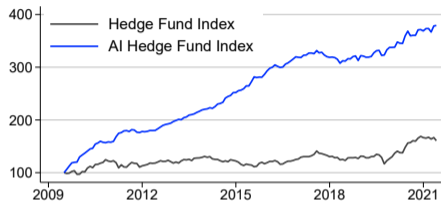
Source: Eurekahedge.

Data from Bloomberg. Returns in USD from 2009-12-31 to 2021-11-30.

Big Data and Machine Learning
What is It? Definitions for this Talk

Applications in Asset Management
More than Risk and Return
Asset Management is Different

Return Prediction and Portfolio Construction
A Framework for Investment Strategies
Real-World Strategies and Funds

Open Issues and Conclusions
Lessons from Industry-Applications
Trade-Offs and Conclusions

Experiences and discussions among practitioners[10] ...

- ▶ Define a narrow problem and keep models simple and robust!
  - ▶ No allocation decisions! Sector-neutral selection worked best.
  - ▶ Not too many features!
  - ▶ Ensembles instead of model selection!

- ▶ Data vs. Models – Where is the competitive advantage?
  - ▶ *Pro data:*   models are open source. Advantage comes from new information.
  - ▶ *Pro model:*   successful strategies that only use standard data (e.g. HQAM).
  - ▶ *Consensus:*   "alternative data" from vendor is no durable advantage.

There are successful applications but the industry is competitive!

---
[10] Source: NextGen Alpha (2018a,b, 2019, 2020a,b, 2021).

Do we need to understand the machine?

I don't know why planets orbit the sun. That doesn't mean I can't predict them.

Jim Simons quoted in Zuckerman (2019, Ch. 8)

*If we have difficulty in **bad times** [...] in **getting people to stick with it**, try to imagine that same bad time for the opaque machine learning factor. [...] People want the new hot, sexy thing and they **don't trust algorithms** at the same time.*

Cliff Asness at the Management Conference '21 (2021, 57:00)

Statistical prediction vs. interpretability and fiduciary duty!

It depends on the strategy[11] ...

	Problems with Machine Learning in Finance	
	Low	**High**
Type of Strategy	▪ High-frequency / fast signals ▪ (Very) high Sharpe ratios ▪ Limited capacity	▪ Low-frequency / slow signals ▪ Lower Sharpe ratios ▪ High capacity
Explicability	▪ **Not required** ▪ Quick performance feedback ▪ Signal switched off if not working	▪ **Very important** ▪ Slow performance feedback ▪ Confidence in signal is important

[11] Own illustration of Winton (2021).

Quants will have to live with less backtests ...

Problem #1: Alternative Data is hard to backtest ...
- ► Data usually not available for entire cross section of firms.
- ► Almost impossible to get long history without survivorship bias.
- ► History of some data physically limited (e.g. social media).

Problem #2: Look-Ahead Bias – also for methodology ...
- ► Classic: simulated trading based on signal that was not available at that time.
- ► For methodology: impossible to trade on predictions of ML in 1980s.
- ► Future performance unlikely as good as backtest.

Conclusion

- ▶ Big Data and Machine Learning are here to stay! It is just getting started ...

  - ▶ Evolutionary: improve established strategies or assist fundamental analysis.
  - ▶ Revolutionary: new funds that are completely managed by machines.

- ▶ Various impacts on the asset management industry:

  - ▶ More efficient capturing of anomalies and known risk-premiums.
  - ▶ Large players have advantage: data and infrastructure are barriers to entry.
  - ▶ Non-quant managers will be forced to augment and adapt their processes.

Huge potential and several promising applications already available!

But investing remains more difficult than image recognition!

Thank you very much!

Website: AgnosticInvesting.com

LinkedIn: Sven Sadlo

Scan Me

References

Bartram, Söhnke M., Jürgen Branke, and Mehrshad Motahari, 2020, Artificial Intelligence in Asset Management, *CFA Institute Research Foundation Literature Review*.

Bartram, Söhnke M., Jürgen Branke, Giuliano De Rossi, and Mehrshad Motahari, 2021, Machine Learning for Active Portfolio Management, *The Journal of Financial Data Science* 3, 9–30.

Goldstein, Itay, Chester Spatt, and Mao Ye, 2021, Big Data in Finance, *National Bureau of Economic Research Working Paper 28615*.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu, 2020, Empirical Asset Pricing via Machine Learning, *The Review of Financial Studies* 33, 2223–2273.

Israel, Ronen, Bryan Kelly, and Tobias Moskowitz, 2020, Can Machines "Learn" Finance?, *Journal Of Investment Management* 18, 23–36.

Laney, Doug, 2001, 3D Data Management: Controlling Data Volume, Velocity, and Variety, *META Group Research Note*.

References (cont.)

Leung, Edward, Harald Lohre, David Mischlich, Yifei Shea, and Maximilian Stroh, 2021, The Promises and Pitfalls of Machine Learning for Predicting Stock Returns, *The Journal of Financial Data Science* 3, 21–50.

Management Conference '21, 2021, Keynote Conversation with Cliff Asness, David Booth, and Eugene F. Fama, *Recording: https://www.youtube.com/watch?v=ltYNbsIhM24*.

Martin, Ian, and Stefan Nagel, 2019, Market Efficiency in the Age of Big Data, *National Bureau of Economic Research Working Paper 26586*.

Nagel, Stefan, 2021, *Machine Learning in Asset Pricing*, 1st Edition (Princeton University Press, Princeton).

NextGen Alpha, 2018a, 3rd NextGen AI Aware Event, *Industry Conference, Recordings: https://nextgen.hvo.events/record-3rd-event/*.

References (cont.)

NextGen Alpha, 2018b, 4th NextGen AI Aware Event, *Industry Conference, Recordings: https://nextgen.hvo.events/record-4th-event/*.

NextGen Alpha, 2019, 5th NextGen AI Aware Event, *Industry Conference, Recordings: https://nextgen.hvo.events/record-5th-event/*.

NextGen Alpha, 2020a, Investment Management Powered by AI, *6th NextGen Event, Industry Conference, Recordings: https://nextgen.hvo.events/record-6th-event/*.

NextGen Alpha, 2020b, Investment Management Powered by AI, *8th NextGen Event, Industry Conference, Recordings: https://nextgen.hvo.events/record-8th-event/*.

NextGen Alpha, 2021, Investment Management Powered by AI, *9th NextGen Event, Industry Conference, Recordings: https://nextgen.hvo.events/record-9th-event/*.

Pedersen, Lasse Heje, 2015, *Efficiently Inefficient: How Smart Money Invests and Market Prices Are Determined*, 1st Edition (Princeton University Press).

References (cont.)

Rasekhschaffe, Keywan Christian, and Robert C. Jones, 2019, Machine Learning for Stock Selection, *Financial Analysts Journal* 75, 70–88.

Sagiroglu, Seref, and Duygu Sinanc, 2013, Big data: A review, in *2013 International Conference on Collaboration Technologies and Systems (CTS)*.

Winton, 2021, Experiment and Observation in Quantitative Investment, *Research*.

Zuckerman, Gregory, 2019, *The Man Who Solved the Market: How Jim Simons Launched the Quant Revolution* (Penguin LCC US).