Financial Data Professional Institute

A Conversation with Joe Simonian

A Machine Learning Approach to Risk Factors: A Case Study Using the Fama-French-Carhart Model

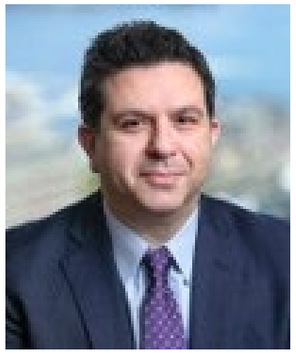
Mehrzad Mahdavi, Executive Director, FDP Institute
Kathy Wilkens, Senior Advisor, FDPI Curriculum
Mirjam Dekker, Project Manager, FDP Institute

www.fdpinstitute.org
February 27, 2020

αutonomous
INVESTMENT TECHNOLOGIES

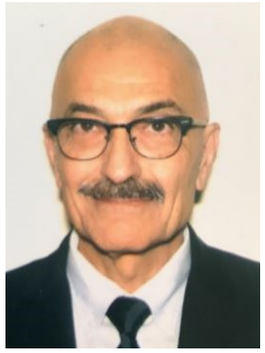
Agenda

- Welcome
- Introductions



Joe Simonian,
Founder and CIO, Autonomous Investment Technologies, LLC;
Co-editor of The Journal of Financial Data Science

Katherine Wilkens
Sr. Curriculum Advisor FDPI



Mehrzad Mahdavi
Executive Director
FDPI

Mirjam Dekker
Project Manager
FDPI

- Fama-French-Carhart Model
- Drawbacks of Traditional Quant Toolkit
- Pros of Applying ML in Finance
- Case Study
- FDP Curriculum

αutonomous
INVESTMENT TECHNOLOGIES

Drawbacks of the Traditional Quant Toolkit

The basis of much of contemporary risk measurement and investment management is based on Ordinary Least Squares (OLS) regression, developed by Carl Friedrich Gauss more than two centuries ago. Nevertheless, it possesses a number of shortcomings that make it a less than ideal framework for analyzing financial data characterized by:

- ❑ Non-linear threshold relationships
- ❑ Variables which exhibit important dependence relationships
- ❑ The interaction between numerical and categorical variables
- ❑ Other statistical and econometric models possess many of these shortcomings (and others), e.g. PCA.

Key takeaway: Investment practitioners current econometric toolkit is inadequate for dealing with the nuances of financial data sets

Joseph Simonian, Ph.D.
Founder and CIO, *AUTONOMOUS INVESTMENT TECHNOLOGIES LLC*
Co-editor of *The Journal of Financial Data Science*

FDP INSTITUTE

Pros of Applying ML to Investing

- Machine Learning frameworks gain much of their problem-solving power from the unique ways in which they process information. For example:
  - ❑ Hierarchical approaches such as Random Forests
  - ❑ Inherent parallel processing such as that found in Neural Networks
  - ❑ Visual/graphical modes of analysis such as that utilized in Spectral Clustering

  The variety of ML methods for processing information is invaluable in finance. For example, the risk management of multi-asset portfolios often involves uncovering and monitoring indirect relationships (e.g. between credit and equity) and detecting the likelihood of tail events that could have cross-asset impact. It often takes more than one approach to discover these relationships.

Key takeaway: Machine learning methods provide the technical means to conduct more nuanced and complete investment analysis

Joseph Simonian, Ph.D.
Founder and CIO, *AUTONOMOUS INVESTMENT TECHNOLOGIES LLC*
Co-editor of *The Journal of Financial Data Science*

αutonomous
INVESTMENT TECHNOLOGIES

Case Study: ML Meets the Fama-French-Carhart Model

Fama-French-Carhart Factor Specification:

$$E[R_s] = r_f + \beta_s^{Mkt}(E[R_{Mkt}] - r_f) + \beta_s^{SMB}E[R_{SMB}] + \beta_s^{HML}E[R_{HML}] + \beta_s^{PR1YR}E[R_{PR1YR}]$$

Mkt: Market Beta

SMB captures Size Effect: Stocks with lower market capitalizations have been found to have higher average returns.

HML captures Value: Stocks with low book-to-market ratio have historically produced positive risk-adjusted returns.

PR1YR captures Momentum: Buying stocks that have had past high returns and shorting stocks that have had past low returns has historically produced higher average returns.

Joseph Simonian, Ph.D.

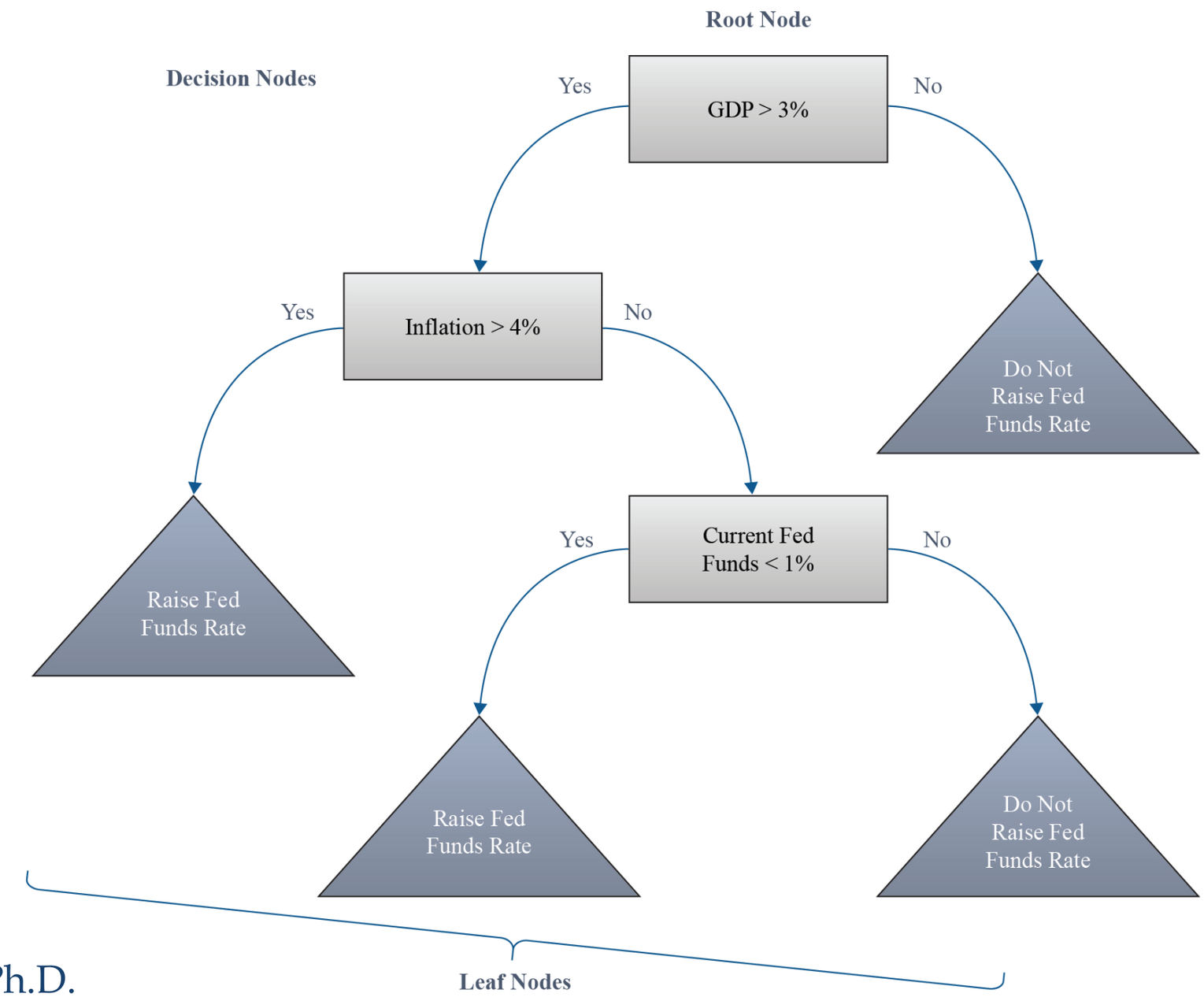
Founder and CIO, *AUTONOMOUS INVESTMENT TECHNOLOGIES LLC*

Co-editor of *The Journal of Financial Data Science*

FDP INSTITUTE

αutonomous
INVESTMENT TECHNOLOGIES

Case Study: Random Forests



Joseph Simonian, Ph.D.
Founder and CIO, *AUTONOMOUS INVESTMENT TECHNOLOGIES LLC*
Co-editor of *The Journal of Financial Data Science*

FDP INSTITUTE™

αutonomous
INVESTMENT TECHNOLOGIES

Case Study: Sample Output

Factor Percentile Returns and Equity Sector Predicted Values (monthly returns, Jan 1991 to Aug 2018)

Percentile	Consumer Discretionary	Rm-Rf	SMB	HML	MoM	Consumer Staples	Rm-Rf	SMB	HML	MoM
10th	–5.2%	–4.7%	–3.3%	–3.0%	–4.7%	–3.3%	–12.4%	–10.8%	–7.6%	–21.2%
25th	–1.7%	–1.9%	–1.8%	–1.5%	–1.4%	–0.9%	–5.3%	–3.6%	–3.6%	–5.4%
50th	1.3%	1.4%	0.3%	0.2%	1.0%	1.0%	2.2%	1.0%	1.0%	1.9%
75th	4.4%	3.7%	2.4%	2.1%	3.3%	2.9%	5.1%	3.4%	3.4%	4.6%
90th	6.7%	5.9%	3.9%	3.9%	5.2%	4.1%	11.4%	22.1%	12.9%	18.3%
R^2	0.96					0.91				
OLS R^2	0.81					0.44				
Percentile	**Energy**	**Rm-Rf**	**SMB**	**HML**	**MoM**	**Financials**	**Rm-Rf**	**SMB**	**HML**	**MoM**
10th	–4.6%	–4.5%	–3.3%	–2.9%	–4.2%	–5.9%	–3.2%	–2.8%	–2.4%	–2.9%
25th	–2.1%	–1.2%	–1.3%	–1.1%	–0.8%	–2.4%	–1.2%	–1.2%	–1.1%	–0.6%
50th	0.8%	0.8%	–0.2%	–0.3%	0.2%	1.6%	1.4%	0.3%	0.1%	0.9%
75th	4.4%	3.8%	2.5%	2.2%	3.4%	4.1%	3.9%	2.5%	2.3%	3.5%
90th	6.5%	7.0%	5.0%	5.5%	7.0%	7.9%	4.8%	3.2%	3.2%	4.1%
R^2	0.90					0.96				
OLS R^2	0.40					0.84				

Source: Simonian et al. "A Machine Learning Approach to Risk Factors: A Case Study Using the Fama-French-Carhart Model", *The Journal of Financial Data Science* 1(1), 2019.

Joseph Simonian, Ph.D.
Founder and CIO, *AUTONOMOUS INVESTMENT TECHNOLOGIES LLC*
Co-editor of *The Journal of Financial Data Science*

FDP INSTITUTE™

αutonomous
INVESTMENT TECHNOLOGIES

Case Study: Sample Output

Sector Rotation Strategy Backtest and Bootstrap Results (Jan 1997 to Aug 2018)

	Annualized Return	Annualized Volatility	PSharpe 0.0	PSharpe 0.1	PSharpe 0.2	PIR 0.0	PIR 0.1	PIR 0.2	Turnover	Total Cumulative Return
Panel A: Out-of-Sample Backtest Results										
Equal-Weight	10.0%	13.8%	99.2%	78.8%	23.4%	N/A	N/A	N/A	0.0%	632%
Unconstrained Strategy	12.9%	14.2%	99.9%	94.9%	53.8%	96.9%	60.3%	10.3%	74.6%	1106%
Constrained Strategy	11.0%	13.6%	99.7%	87.3%	35.0%	97.8%	66.1%	13.6%	22.4%	773%
Panel B: Bootstrap Performance Results										
Equal-Weight	11.0%	13.4%	96.7%	78.9%	43.0%	N/A	N/A	N/A	0.0%	982%
Unconstrained Strategy	12.2%	14.4%	97.6%	82.2%	45.8%	66.7%	25.7%	4.2%	85.4%	1202%
Constrained Strategy	11.4%	13.4%	97.3%	81.3%	45.9%	66.7%	25.8%	4.3%	25.6%	1041%

Source: Simonian et al. "A Machine Learning Approach to Risk Factors: A Case Study Using the Fama-French-Carhart Model", *The Journal of Financial Data Science* 1(1), 2019.

Joseph Simonian, Ph.D.

Founder and CIO, *AUTONOMOUS INVESTMENT TECHNOLOGIES LLC*

Co-editor of *The Journal of Financial Data Science*

FDP INSTITUTE™

FDP Curriculum & Ethics

Sample of the **Reading(s):**

8.2 Phoon, K. and F. Koh. (2018). Robo-Advisors and Wealth Management. *Journal of Alternative Investments*, 20(3), 79-94. DOI: https://doi.org/10.3905/jai.2018.20.3.079

8.3 Monk, A., M. Prins, and D. Rook. (2019). Rethinking Alternative Data in Institutional Investment. *Journal of Financial Data Science*, 1(1), 14-31. DOI: https://doi.org/10.3905/jfds.2019.1.1.014

8.4 Simonian, J., C. Wu, D. Itano and V. Narayanan. (2019). A Machine Learning Approach to Risk Factors: A Case Study Using the Fama–French–Carhart Model. *Journal of Financial Data Science*, 1(1), 32-44. DOI: https://doi.org/10.3905/jfds.2019.1.032

Sample Keywords *(of the Simonian reading):*

Keywords

Factors (p. 32)
Linear (p. 32)
Nonlinear (p. 32)
Random forest (p. 33)
Supervised (p. 34)
Unsupervised (p. 34)
Root node (p. 34)
Decision node (p. 34)
Terminal node (p. 34)
CART (p. 34)
Binary recursive partitioning (p. 34)
Bagging (p. 34)
Out-of-bag data (p. 34)
Feature importance (p. 35)
Mean decrease accuracy (p. 34)
Fama-French-Carhart (p. 37)
Probabilistic Sharpe ratio (p. 42)

Source: FDP Institute Study guide March 2020 Exam

FDP INSTITUTE

FDP Curriculum

Sample Learning Objectives

8.4.1 Applications of random forest regression algorithm to factor models

For example:

A. Discuss two shortcomings of parametric nonlinear factor models that are developed to address shortcomings of linear models.

B. Discuss the ability of random forest algorithm to overcome one shortcoming of linear models.

C. Discuss the ability of random forest algorithm to overcome one shortcoming of parametric nonlinear models.

D. List 4 components of the decision tree when applied to the regression problem of factor models.

E. Describe how bagging is used in an ensemble of decision trees (random forest).

F. Calculate the predicted value of an dependent (response) variable given a set of predictor values and the outputs of a binary regression decision tree algorithm.

G. Describe the role of out-of-bag observations in a random forest algorithm.

H. Discuss mean decrease accuracy approach to estimating feature importance in a random forest algorithm.

I. Recognize and apply the probabilistic Sharpe ratio.

Sample Questions:

According to "A Machine Learning Approach to Risk Factors," by Simonian et al., what is one of the shortcomings of non-linear factor models in comparison to linear factor models?

a) The impacts of correlated factors are difficult to measure
b) The parameter estimates cannot always be derived analytically
c) The number of factors that must be used is relatively large

Source: FDP Institute Study guide March 2020 Exam

Answer b
Source: LO 8.4.1, Reading 8.4, pp 32-33

Q & A

Kind reminders of upcoming webinars as we go through the Q & A.

Add you questions in the chat room please.

CAIA ASSOCIATION FDP INSTITUTE

WEBINAR SERIES

A Conversation With...

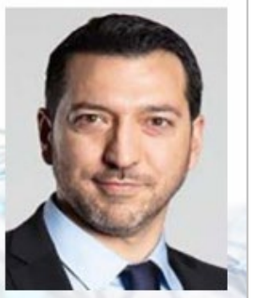
Tony Guida

Executive Director, Sr. Quant Research

Ram Active Investment

March 5, 2020

10 am EST



CAIA ASSOCIATION FDP INSTITUTE

WEBINAR SERIES

A Conversation With...

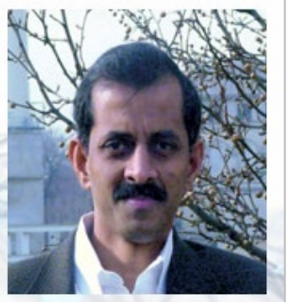
Ganesh Mani

Adj. Faculty Carnegie Mellon

"Data Supply Chain Mgmt."

March 10, 2020

1pm EST



CAIA ASSOCIATION FDP INSTITUTE

WEBINAR SERIES

A Conversation With...

Rick Roche, CAIA

Man. Dir. Little Harbor Advisors

"Evolution of Machine Learning in Investment Mgmt."

March 17 - 11am EST

CAIA ASSOCIATION FDP INSTITUTE

WEBINAR SERIES

A Conversation With...

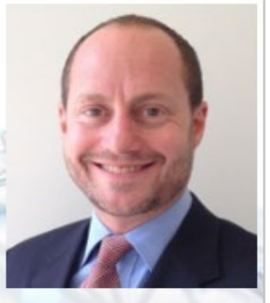
Michael Oliver Weinberg

Managing Dir. , Head of Hedge Funds & Alt. Alpha

APG

Autonomous Learning Investment Strategies

March 25th – 4pm EST



CAIA ASSOCIATION FDP INSTITUTE

WEBINAR SERIES

A Conversation With...

George Mussalli, CFA
PanAgora

Mike Chen, Ph.D.
PanAgora

"An integrative Approach to Quantitative ESG Investing"

April 1, 2020 @ 12noon

FDP INSTITUTE

In Closing

- The Next FDP Exam: March 16 –April 4, 2020
- Registration is open
- For a recent candidate webinar go to www.fdpinstitute.org/webinars

Learn more about the FDP Institute at
www.fdpinstitute.org

FDP INSTITUTE