Opportunities and challenges of using machine learning $\mathbb{ML}$ in quantitative wealth and investment management $\mathbb{QWIM}$

Cristian Homescu

March 2021

Why machine learning $\mathbb{ML}$ for quantitative investment and wealth management $\mathbb{QWIM}$?

$\mathbb{ML}$: relevant tools and sources of information

$\mathbb{ML}$ in $\mathbb{QWIM}$: classification and pattern recognition

$\mathbb{ML}$ in $\mathbb{QWIM}$: network analysis and clustering

$\mathbb{ML}$ in $\mathbb{QWIM}$: forecasting of financial time series

$\mathbb{ML}$ in $\mathbb{QWIM}$: reinforcement learning $\mathbb{RL}$

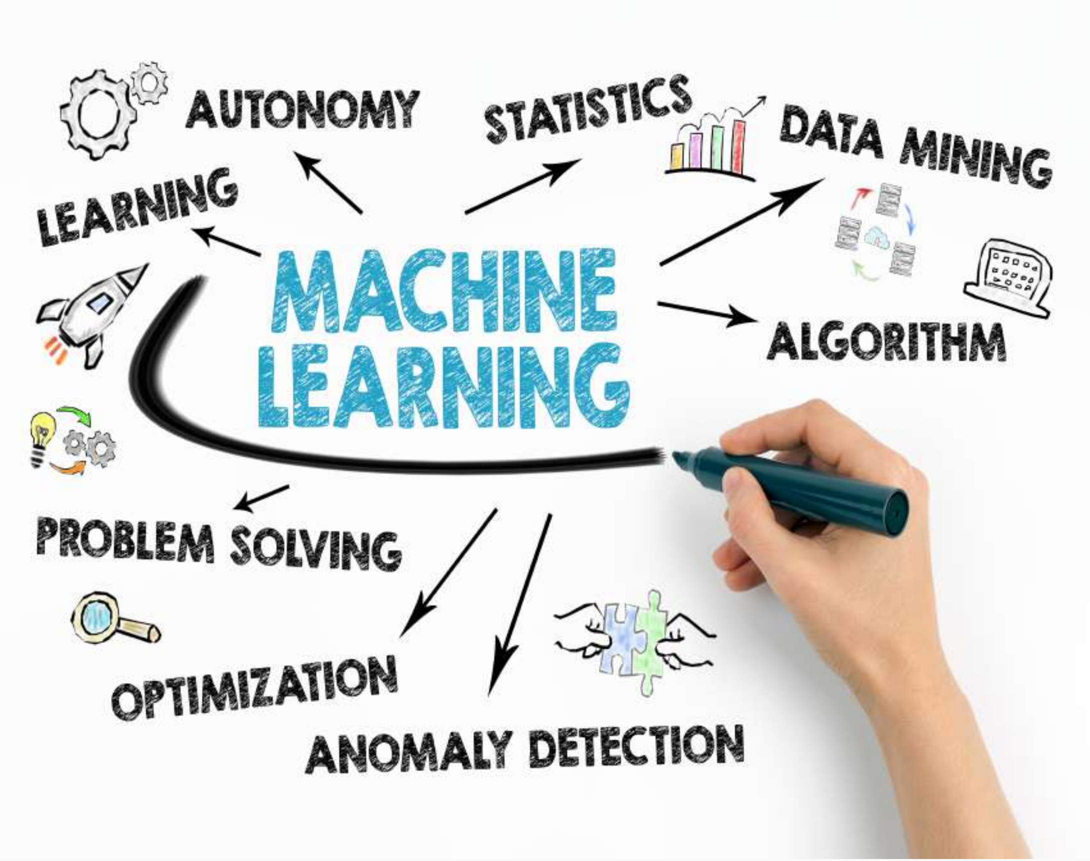
$\mathbb{ML}$ in $\mathbb{QWIM}$: importance of market states and regimes

$\mathbb{ML}$ in $\mathbb{QWIM}$: other practical applications

$\mathbb{ML}$ in $\mathbb{QWIM}$: practical challenges

The views expressed here are those of the author and do not necessarily represent or reflect the views of neither Bank of America nor of its Global Wealth and Investment Management (GWIM) division.

The examples are for illustrative purpose only and do not constitute investment advice.


AUTONOMY
STATISTICS
DATA MINING
LEARNING
MACHINE LEARNING
ALGORITHM
PROBLEM SOLVING
OPTIMIZATION
ANOMALY DETECTION

Overview of the presentation

- Increasingly apparent that Financial Machine Learning is a rather specialized area instead of just a combination of standard Machine Learning and Financial Data.
- What is hype and what is reality when applying machine learning $\mathbb{ML}$ to quantitative investment and wealth management $\mathbb{QWIM}$?
- Presentation delves into what appears to work well (or not so well) for $\mathbb{ML}$ in various areas of $\mathbb{QWIM}$, and describes practical challenges
- Area of classification and pattern recognition:
  - classification and partition of the investment universe
  - investing based on alternative data
  - text analysis of company and regulatory documents
  - sentiment analysis of news and social media
  - ESG (Environmental, Social and Governance)
  - investing fund decomposition, inference and replication

Overview of the presentation (cont.)

- Area of network analysis and clustering:
  - clustering-based portfolio optimization
  - network-based portfolio optimization
  - analysis of interconnectedness risk
  - network effects on investment portfolios.
- Area of time series forecasting:
  - forecasting of financial time series
  - empirical asset pricing
- Area of reinforcement learning:
  - pricing and hedging of financial derivatives
  - optimal dynamic trading strategies
  - portfolio allocation
  - goals-based investing

Overview of the presentation (cont.)

- Other practical applications:
  - synthetic financial data generation
  - testing investment strategies and portfolios
  - factor-based investment strategies
  - nowcasting
  - incorporating market states and regimes into investment portfolios.
- Practical challenges include
  - lack of sufficient data
  - need to satisfy privacy, fairness and regulatory requirements
  - model overfitting
  - causality
  - explainability and interpretability
  - hyperparameter tuning

Why machine learning $\mathbb{ML}$ for quantitative investment and wealth management $\mathbb{QWIM}$?

$\mathbb{ML}$ is everywhere, even to assess wine quality

- Bouri et al. (2018), Le Fur and Outreville (2019): wine as an investment asset class is based on wine quality certification given by human experts, a process both labor intensive and time consuming
- Aich et al. (2019), Gupta (2018): $\mathbb{ML}$ able to assess quality of wines using attributes identified through feature selection
- Accuracy ranging from 92% to 98%

$\mathbb{ML}$ has strong potential in $\mathbb{QWIM}$

- Big data and machine learning have the potential to profoundly change the investment landscape, due to
  - ongoing expansion in amount and type of available data
  - performance and capabilities of $\mathbb{ML}$ packages and platforms
  - computational power and data storage $\Uparrow$, while costs $\Downarrow$
  - significant improvements in $\mathbb{ML}$: methods and software
- Lopez de Prado (2020a), Dixon et al. (2020), Bartram et al. (2020), Jurczenko et al. (2020): considerable advantages of $\mathbb{ML}$ in $\mathbb{QWIM}$: incorporate nonlinearity, pattern recognition, networks and deeper interactions
- Lopez de Prado (2019): *"economic systems exhibit a degree of complexity that is beyond the grasp of classical statistical tools"*
- $\mathbb{AI}$ / $\mathbb{ML}$ already used successfully in $\mathbb{QWIM}$: Guida (2019), Lopez de Prado (2018), Kolanovic and Krishnamachari (2017), Sirotyuk (2019), Lopez de Prado (2020a), Dixon et al. (2020), Jurczenko et al. (2020), Jansen (2020), Bartram et al. (2020)
- $\mathbb{AI}$ / $\mathbb{ML}$ can provide powerful new insights using classical investment data, alternative data, or a combination of both

$\mathbb{ML}$ has strong potential in $\mathbb{QWIM}$ (cont.)

- Efron and Hastie (2016): "*Our new computation-rich environment has unplugged the mathematical bottleneck, giving us a more realistic, flexible, and far-reaching body of statistical techniques*"
- Lopez de Prado (2020a), Jurczenko et al. (2020): $\mathbb{ML}$ for asset managers
  - $\mathbb{ML}$ can help to discover economic and financial theories; complement rather than replace classical methods
  - Some of $\mathbb{ML}$ strengths include:
    - focus on out-of-sample predictability over variance adjudication;
    - usage of computational methods to avoid relying on (potentially unrealistic) assumptions
    - ability to "learn" complex specifications, including non-linear, hierarchical and non-continuous interaction effects in a high-dimensional space
    - ability to disentangle the variable search from the specification search, robust to multicollinearity and other substitution effects

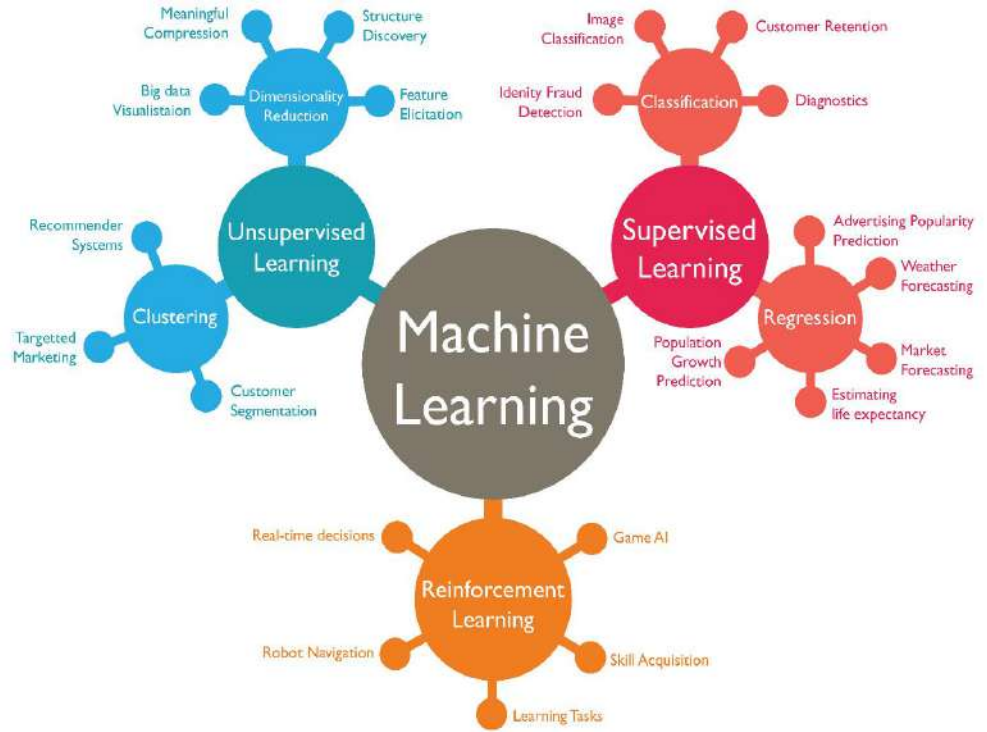
However ...

- To extract significant value from $\mathbb{QWIM}$ datasets, standard $\mathbb{ML}$ may need to be enhanced (from modeling, numerical and computational perspectives) and augmented with domain expert knowledge
- Lopez de Prado (2018): Financial Machine Learning is rather specialized

Financial $\mathbb{ML}$ $\neq$ $\mathbb{ML}$ Algorithms + Financial Data

- Heaton (2018): "*financial market data is unlike the data that machine learning works well on in computer vision, speech recognition, and natural language processing*."
- Thus significant challenges when using $\mathbb{AI}$ /$\mathbb{ML}$ in $\mathbb{QWIM}$
  - Datasets may not be sufficient (quantity and/or quality) for $\mathbb{ML}$
  - Financial datasets have distinctive stylized characteristics
  - Model overfitting and data mining risks may be higher for $\mathbb{ML}$
  - Forecasting $\mathbb{QWIM}$ time series may need more than standard $\mathbb{ML}$
  - Stakeholders and regulators require interpretability and/or explainability for $\mathbb{ML}$

Visual categorization of machine learning algorithms



Source: Jha (2017)

Important ML applications in QWIM

- Classification and pattern recognition can be used for
  - inclusion of alternative data into investment process
  - fund classification, inference and decomposition
  - topic and sentiment analysis
  - risk detection from corporate documents
- Reinforcement learning can be used for
  - goals-based investing
  - construction and risk management of trading and portfolios
  - retirement and financial planning
- Network analysis and clustering can be used for
  - portfolio optimization
  - analysis and improvement of portfolio diversification
  - clustering of financial time series
  - examination of network-related risks

Important $\mathbb{ML}$ applications in $\mathbb{QWIM}$ (cont.)

- Forecasting and prediction can be used for
  - Short-term, medium-term and long-term forecasting of returns and risks for asset classes and investment vehicles
  - Forecasting of alternative risk premia
  - Prediction of market regimes and structural breaks
  - Forecasting fundamentals and earnings of individual companies
- Other practical applications in $\mathbb{QWIM}$ include
  - regime-based investing
  - testing and analysis of trading and investment strategies
  - scenario generation and data augmentation for financial time series
  - ESG-based investing
  - factor-based quantitative investing
  - due diligence and document analysis
  - nowcasting
  - compliance with internal rules and external regulations

$\mathbb{ML}$: relevant tools and sources of information

- Explosive growth of information related to $\mathbb{ML}$: books, articles, conference presentations, preprints and working papers, Python and R packages, GitHub and GitLab repositories
- One can efficiently find relevant research works through specialized online avenues: ConnectedPapers.com, Dimensions.ai, PapersWithCode, SSRN, arXiv, arXiv Sanity, SemanticScholar, papertalk.org
- Overviews and best practices: Lopez de Prado (2020a), Kolanovic and Krishnamachari (2017), Burkov (2019), Lopez de Prado (2018), Ryll and Seidens (2019), Bloch (2020b), Bloch (2020a), Dixon and Halperin (2019), Sammut and Webb (2017), Hutter et al. (2019), Malik (2020), Goan and Fookes (2020), Raschka et al. (2020), Krohn et al. (2019), Burkov (2019), Boehmke and Greenwell (2019), Chollet (2017), Lapan (2020), Jurczenko et al. (2020), Dixon et al. (2020)
- Machine learning platforms and frameworks
  - TensorFlow, PyTorch, Theano, PlaidML, Keras, Microsoft Cognitive Toolkit, mlr3, MXNet, Caffe, Deeplearning4j, Chainer, SpaCy
  - H2O.ai, DataRobot, OneClick.ai, DarwinAI, Dataiku, Valohai
  - Google TFX, AWS SageMaker, IBM Watson ML, Microsoft Azure ML
- Small data and machine learning: Greco et al. (2019), Li et al. (2018), Qi and Luo (2019), Brigato and Iocchi (2020), Yao et al. (2020a)

Major topics in machine learning

- Deep learning $\mathbb{DL}$: Goodfellow et al. (2016), Nakagawa et al. (2019b), Hatcher and Yu (2018), Polson and Sokolov (2018b), Zhang et al. (2018b), Alonso et al. (2019), Polson and Sokolov (2018a), Higham and Higham (2018), Aguilera et al. (2020), Bacciu et al. (2020), Li et al. (2020a), Ozbayoglu et al. (2020), Raghu and Schmidt (2020), Zhang et al. (2020a), Thompson et al. (2020), Samek et al. (2020), Xie et al. (2020), Ren et al. (2020)
- Natural Language Processing: Aggarwal (2018), Kamath et al. (2019), McMahan and Rao (2019), Lane et al. (2019), and Eisenstein (2019), Qiu et al. (2020), Torfi et al. (2020), Ferrario and Naegelin (2020), Chen et al. (2020a)
- Reinforcement learning $\mathbb{RL}$ (also Deep $\mathbb{RL}$ and inverse $\mathbb{RL}$): Dong et al. (2019), Francois-Lavet et al. (2018), Nachum et al. (2019), Sutton and Barto (2018), Zhang et al. (2018a), Zhao et al. (2019), Xing (2019), Laskin et al. (2020), Lapan (2020)
- Probabilistic graphical models: Benhamou (2018), Denev et al. (2018), Denev and Mutnikas (2016a), Gao et al. (2018), Ma and Zhang (2019)
- Transfer learning: Ada et al. (2019), Fawaz et al. (2018b), He et al. (2019a), Krishna and Menzies (2018), Sun et al. (2019), Nakagawa et al. (2019a)
- Tuning hyperparameters: Cho and Hegde (2019), Yu and Zhu (2020), Feurer and Hutter (2019), Katakami et al. (2019), Kim and Cho (2019), Klein et al. (2019), Sandru and David (2019), Sanders and Giraud-Carrier (2017), Tu and Nair (2018), van Rijn and Hutter (2018), Yang and Shami (2020), Yu and Zhu (2020)

Major topics in machine learning (cont.)

- Generative Adversarial Networks (GANs): Saxena and Cao (2020), Jabbar et al. (2020), Koochali et al. (2020), Ni et al. (2020a), Yoon et al. (2019), Vieira (2020)
- Variational Autoencoders: Yu (2020), Charte et al. (2018), Imokoyende (2019), Kingma and Welling (2019), Rocca and Rocca (2019), Yacoby et al. (2020)
- Data augmentation: Fawaz et al. (2018a), Hoffmann et al. (2018), Kuchnik and Smith (2018), Taylor and Nitschke (2017), Tran et al. (2017), Volpi et al. (2018), He et al. (2019b), Wang et al. (2019a), Xie et al. (2019), Wen et al. (2020), Fang and Lin (2020), Saxena and Cao (2020), Laskin et al. (2020)
- Feature selection: Cai et al. (2018), Nogueira et al. (2018), Sreevani and Murthy (2017), Sreevani and Murthy (2017), Stanczyk and Jain (2015), Bolon-Canedo and Alonso-Betanzos (2019), Bugata and Drotar (2019), Draminski and Koronacki (2018), Lu et al. (2018),Piliszek et al. (2019), Yan and Bien (2018), Barandas et al. (2020)
- Overfitting and how to control it: Hastie et al. (2011), James et al. (2015), Aparicio and Lopez de Prado (2018), Bailey et al. (2017), Philipp and Carbonell (2018), Suhonen et al. (2017), Zhang et al. (2018a), Koshiyama and Firoozye (2019), Rasekhschaffe and Jones (2019), Lopez de Prado (2020b)
- Anomaly detection: Bulusu et al. (2020), Braei and Wagner (2020), Bashchenko and Marchal (2020), Das et al. (2019c), Lee et al. (2020), Wu (2016), Wu and Keogh (2020), Jacob et al. (2020b), Ruff et al. (2020)

Major topics in machine learning (cont.)

- Network analysis: Bramson and Vandermarliere (2016), Das et al. (2019a), Fortunato and Hric (2016), Iori and Mantegna (2018), Jeub et al. (2015), Lacasa et al. (2015), Lawyer (2015), Runge (2018), Zanin et al. (2016), Marti et al. (2020)
- Clustering: Maharaj et al. (2019), Cai et al. (2016), Cavallo and Demiralp (2018), Durante et al. (2013), Kauffmann et al. (2019), Khaleghi et al. (2016), Murtagh and Contreras (2012), Murtagh and Contreras (2017), Patel and Thakral (2016), Adolfsson et al. (2019), Javed et al. (2020), Sartorio and Fonseca (2020)
- Interpretability: Liu et al. (2018), Kim and Doshi-Velez (2017), Poursabzi-Sangdeh et al. (2018), Zhang and Zhu (2018), Avigdor-Elgrabli et al. (2018), Du et al. (2018), Gilpin et al. (2018), Hall et al. (2018), Mohseni et al. (2018), Murdoch et al. (2019), Schmidt and Biessmann (2019), Friedler et al. (2019), Ghorbani et al. (2019), Hazard et al. (2019), Ish-Horowicz et al. (2019), Ahuja et al. (2018), Bachoc et al. (2018), Brown and Petrik (2018), Rudin (2019), Choi et al. (2020), Caruana et al. (2020), Molnar et al. (2020)
- Explainability: Arik and Pfister (2019), Adadi and Berrada (2018), Amarasinghe et al. (2018), Choo and Liu (2018), Dosilovic et al. (2018), Horel and Giesecke (2018), Preece (2018), Roscher et al. (2019), Vasic et al. (2019), Ahern et al. (2019), Arras et al. (2019), Gade et al. (2019), Samek and Muller (2019), Singh and Anand (2019), Vaughan et al. (2018), Moore et al. (2019), Ignatiev (2020), Samek et al. (2020), Shi et al. (2020), Xie et al. (2020), Puiutta and Veith (2020), Moreira et al. (2020), Moraffah et al. (2020)

$\mathbb{ML}$ in $\mathbb{QWIM}$: classification and pattern recognition

What type of alternative data is available in $\mathbb{QWIM}$?

- Ekster and Kolm (2020): usage of alternative data in $\mathbb{QWIM}$
- Kolanovic and Krishnamachari (2017) and Guida (2019):
  - Alternative data can be generated by:
    - individuals (social media, news and reviews, web searches)
    - business processes (e-commerce, credit card transactions, company data)
    - government agencies (regulations, rules, guidelines, etc.)
    - machines/sensors (satellite imagery, radar, traffic, etc.)
  - Alternative data can incorporate many categories of attributes:
    - asset class: equity, fixed income, FX, commodity
    - investment style: macro, sectors, risk indicators
    - processing type: raw, processed, trading signal, research
    - quality: historical, outliers, missing values, transparency
    - technical: format, latency, frequency, legal

Visual categorization of alternative data

Source: Lipus and Smith (2019)

Classification and partition of investment universe

- Challenges of partition into asset class categories: a) if partition is "too coarse", portfolio may be missing desired asset/subasset classes versus b) if partition is "too fine", portfolio may end up with very small allocations, while costs would increase (trading, compliance, etc.)
- Lamponi (2015): data-driven categorization of investable assets
- Ma et al. (2019): unsupervised $\mathbb{ML}$ to identify number of most relevant risk-axes and corresponding asset categories, then supervised $\mathbb{ML}$ to identifymost important defining characteristics of an asset category
- Garvey and Madhavan (2019): model output (geographical footprint) identified by $\mathbb{ML}$ may differ significantly from usual groupings
- Kakushadze and Yu (2017a), Kakushadze and Yu (2017b): $\mathbb{ML}$ when industry and sector classification not very clear, or needs to adjust rapidly
- Yang et al. (2016), Kee (2019): automated, text-based industry and sector classification reflecting constant changes in companies' scope
- Fang et al. (2020): GICS sector categorization can be successfully reconstructed from historical fundamental data using $\mathbb{ML}$

Classification and partition of investment universe (cont.)

- Gao et al. (2020): text-based industry classification using word embedding schemes and clustering algorithms
- Mehta et al. (2020): industry categorization system using $\mathbb{ML}$ is reproducible and may be as good (or better) than categorization based on human curation
- Kilburn (2020b): many asset managers no longer convinced of asset class labels used for decades; use clustering and network analysis rather than rely on traditional information such as their sector, geography, issuer or credit ratings
- Kilburn (2019b): $\mathbb{ML}$ can group assets better and predict regime shifts
- Husmann et al. (2020): t-distributed stochastic neighbor embedding (t-SNE) and spectral clustering for company classification
- Massaro et al. (2020): $\mathbb{ML}$ delivers very reliable predictions company classification based on sector of economic activity

Text analysis of company and regulatory documents

- Loughran and McDonald (2020): survey of textual analysis in finance
- Feuerriegel and Gordon (2018): text-based and news-based $\mathbb{ML}$ improves forecasts of stock indices and interpretability
- DiPietro (2019), Fleiss et al. (2020): $\mathbb{ML}$ in combination with SEC 13f filling data found to be effective in terms of portfolio cloning
- Das et al. (2019b): NLP to parse corporate email content and news to assess factors that predict escalating risk or potential firm malaise before they manifest in observable data and financial outcomes
- Klevak et al. (2019): scoring of earnings conference call transcripts based on tone (sentiment) and specific events delivers a signal incrementally additive to earnings surprises and short-term returns around earnings announcement
- Ravula (2021a), Ravula (2021b): text analysis in financial disclosures and and bankruptcy prediction using disclosure text features

Text analysis of company and regulatory documents (cont.)

- Agarwal (2020): systematic textual analysis of company annual reports delivers sustainable returns orthogonal to market returns
- Cohen et al. (2019): changes to 10-Ks predict future earnings, profitability, future news announcements, and even future firm-level bankruptcies
- Amel-Zadeh et al. (2020): $\mathbb{ML}$ capable of forecasting sign and magnitude of abnormal stock returns around earnings announcements based on financial statement data alone
- Luo (2020): NLP applied to company management presentations
  - to analyze readability index and language complexity
  - to quantify executive personalities
- NLP delivers alpha factors uncorrelated with traditional factors, offering significant diversification benefits

Sentiment analysis of news and social media

- Naumer and Yurtoglu (2020), Houlihan and Creamer (2017), Liew and Budavari (2017): Social media and news sentiments have significant power in explaining time series variation in stock returns
- Frydman et al. (2021): mechanism describing how market sentiment drives forecasts of stock returns
- Huang et al. (2018): investor sentiment relevant for stocks, bonds, commodities, currencies, housing
- Audrino et al. (2020): sentiment and attention from social media, news articles, information consumption, and search engine data have significant impact on stock market volatility
- Coqueret (2020b): quantifies impact of stock-specific news sentiment on future financial returns
- Tran (2021): bridging model that connects risk-based factor models to sentiment models
- Ravula (2021b): market sentiment analysis using Valence Aware Dictionary and sEntiment Reasoner

Sentiment analysis of news and social media (cont.)

- Elagamy et al. (2018): text mining combined with Random Forest may determine early warning indicators for stock market crisis
- Kadous et al. (2017): investment advice provided on social media is influential even when it has little predictive value.
- Creamer (2015): since Black-Litterman model is meant to incorporate investor views, why not use sentiment from news, social media and corporate social network as source of investors' expectations?
- Martin-Utrera (2020), Hilliard et al. (2020): market sentiment relevant for portfolio optimization
- Luo (2020), Adammer and Schussler (2020): information embedded in news a) has predictive power and b) is not captured by established economic predictors
- Upreti et al. (2019), Man et al. (2019), Mohammad (2020): surveys of $\mathbb{ML}$-based financial news analytics and sentiment analysis

Environmental, Social and Governance ESG Investing

- BNP Paribas (2019): More than 75% of asset owners and 60% of asset managers hold $\geq 25\%$ of their investments in funds incorporating ESG
- Walker (2019): "*There are issues with the fundamental analyst driven nature of ESG factors in investment process. Data can be too infrequent, measuring events are sometimes too extreme, and capturing events can be too late*".
- Nauman (2019): $\mathbb{ML}$ can go through vast reams of ESG data in a sector where information comes from many sources and is often presented in incompatible ways due to lack of reporting regulations.
- Laidlaw (2019): "*AI is really the catalyst that allows us to find very fine-grain nuggets of information in massive unstructured data*"
- de Franco (2019b), De Franco et al. (2020): $\mathbb{ML}$ can identify patterns between ESG profiles and financial performances for companies in a large investment universe
- Guo et al. (2020b): pipeline of ESG news extraction, news representations, and Bayesian inference of deep learning models

Environmental, Social and Governance ESG Investing (cont.)

- Antoncic et al. (2020): combines $\mathbb{ML}$ and a sustainability lens based on United Nations Sustainable Development Goals
- Procedure delivers scores, rankings, ratings and benchmarks
- Craik (2019): How $\mathbb{ML}$ can help find ESG opportunities
- However, ESG investing may have significant practical challenges: Cornell (2020), Billio et al. (2020), Bahra and Thukral (2020), Gidwani (2020), Dimson et al. (2020), Harper (2020), Chen and Mussalli (2020), de Franco (2019a), Cornell and Damodaran (2020)
- $\mathbb{ML}$ can resolve some of these challenges
- Lanza et al. (2020): overcoming current inconsistencies in ESG scores by using $\mathbb{ML}$ to identify indicators that contribute better to portfolios
- Mannix (2020): Investors use $\mathbb{ML}$ to incorporate raw data over ratings in ESG alpha hunts
- Sokolov et al. (2021): DL and NLP for automated ESG scoring

Fund decomposition, inference and replication

- Question: *Given only the historic net asset value of a mutual fund, which members of some universe of stocks are held by the fund?*
- Discovering an exact solution is combinatorially intractable
- Byrd et al. (2019): use $\mathbb{ML}$ to produce a computationally efficient inference, and identify ETF constituents with accuracy of 88% to 98%
- Duanmu et al. (2020), Giuzio et al. (2018), Simonian and Wu (2019a): $\mathbb{ML}$ successfully used for replication of less liquid funds (such as hedge funds or private equity funds) using liquid investment vehicles
- Mannix (2019), Wigglesworth (2019): firms using $\mathbb{ML}$ to trade public-market stocks to replicate risk and returns of private equity PE funds.
- Some mimic PE's purported bet on buying small, cheap companies; some bet on sectors; others seek to replicate how PE firms think
- Li and Rossi (2021): $\mathbb{ML}$ to select mutual funds from stocks they hold

$\mathbb{ML}$ in $\mathbb{QWIM}$: network analysis and clustering

Beware of correlations which are not consistent with intuition

- In many cases, combination of increasing data, computational power computers and $\mathbb{ML}$ may not deliver better results in $\mathbb{QWIM}$, especially when decision making relies on data-based spurious correlations rather than causality
- Vigen (2019): many examples of such data-based spurious correlations
  - Divorce rate in Maine has 99% correlation with per capita consumption of margarin
  - Per capita consumption of mozzarella cheese has 96% correlation with civil engineering doctorates awarded in US
  - Per capita consumption of chicken has 90% correlation with total US crude oil imports
- Example in $\mathbb{QWIM}$ Laurinaityte et al. (2019): population growth of captive Asian elephants explains cross-section of expected returns of usually sorted portfolios with $R^2 = 0.91$ and $tStat = 2.93$ for market price of risk.
- *Does it mean that number of captive elephants is the new outstanding factor in empirical asset pricing?*

Clustering-based portfolio optimization

- Practical problem: correlation matrix lacks the notion of hierarchy.
- Thus all investments are potential substitutes of each other
- Since investment vehicles could be grouped in terms of industry, size, liquidity, region, some investments are "closer" substitutes of one another, while other investments are "complementary" to one another.
- Then it is better if similar investments can be placed together, while dissimilar investments can be placed far apart.
- This can be accomplished by leveraging network analysis and graph theory to better describe the dependencies among investment vehicles.
- Clustering-based hierarchical structure has two desirable features:
    - portfolio weights primarily are rebalanced among peers at various hierarchical levels
    - weights are distributed top-down, consistent with how many asset managers construct their portfolios, e.g. from asset class to sectors to individual securities

Clustering-based portfolio optimization (cont.)

- Durante et al. (2015) construction of portfolios diversified in their tail behavior by selecting only a single asset in each cluster
- Dose and Cincotti (2005): enhanced index tracking based on clusters of financial time series
- Tola et al. (2008): clustering improves ratio between predicted and realized portfolio risk.
- Dees et al. (2019): portfolio optimization based on graph cuts/ topology
- Durante et al. (2013): since clusters tend to be comonotone in their extreme low values, diversify over these clusters to avoid contagion during risky scenarios
- Chang et al. (2016): portfolio selection based on Affinity Propagation clustering
- Lemieux et al. (2014): clustering effects on portfolio formation and risk
- Hierarchical clustering (discussed next) refers to formation of a recursive clustering, suggested by the data, not defined a priori

Clustering-based portfolio optimization (cont.)

- Zhang and Maringer (2011) and Zhang and Maringer (2010): clustering asset allocation scheme has better performance
- Lopez de Prado (2016): Hierarchical Risk Parity HRP groups similar investments into clusters, based on a proper dissimilarity metric
- Lohre et al. (2020): use HRP to account for tail dependencies in multi-asset multi-factor allocations
- Raffinot (2017): hierarchical clustering-based asset allocation HCAA is more robust and diversified, and achieve statistically better risk-adjusted performances
- Raffinot (2018): Hierarchical Equal Risk Contribution portfolio merges and enhances HRP and HCAA
- Molyboga (2020): modified HRP to improve diversification within and across clusters
- Leon et al. (2017b) and Leon et al. (2017a): Risk-Adjusted Portfolio Construction based on clustering

Clustering-based portfolio optimization (cont.)

- Jain and Jain (2019): use clustering to account for misspecification of covariance matrix, and to investigate when $\mathbb{ML}$-based portfolios outperform traditional risk-based portfolios
- Duarte and De Castro (2020): framework for asset allocation based on partitional clustering algorithms (intragroup and intergroup)
- Puerto et al. (2020): measuring effect (within portfolio optimization) of clustering on selected assets with respect to non-selected ones
- Lopez de Prado (2020a): practical clustering for portfolio optimization
- Begusic and Kostanjcar (2019): cluster-based shrinkage of correlation matrices for portfolio optimization
- Kolrep et al. (2020): intuition and nature of hierarchical clustering in context of multi-asset multi-factor investing
- Jaeger et al. (2021): Interpretable ML and clustering for diversified portfolio construction

Network-based portfolio optimization

- Networks enable practical usage of high / low centrality concepts
  - significant interconnectedness risk (tail events propagate more quickly) due to assets with high centrality scores
  - "peripheral assets" carry relatively less interconnectedness risk
- Pozzi et al. (2013): investments in “peripheral” stocks (located in poorly connected regions of financial networks) better diversified
- Clemente et al. (2019), Clemente et al. (2021): network-based portfolio consistently outperforms standard portfolios out-of-sample
- Kaya (2015): vol strategy augmented by eccentricity centrality measure
- Zhao et al. (2018), Clemente et al. (2018): optimal portfolios obtained through a network-based approach are composed mainly of peripheral assets
- Li et al. (2019): peripherality in a network used as indicator to identify optimal assets
- Peralta and Zareei (2016): negative relationship between centrality of stocks and their risk-adjusted returns.

Network-based portfolio optimization (cont.)

- Konstantinov et al. (2020): compared factor and asset allocation portfolios using both traditional and network-based approaches
- Demonstrate advantages of graph theory for portfolio management, and dynamic nature of assets and factors with their importance scores
- Centrality scores help to identify crowded exposure and build diversified allocations
- Vyrost et al. (2019): construct financial networks in which nodes are represented by assets and where edges are based on long-run correlations
- Proposed adjustments to portfolio strategies utilize centralization measures from financial networks, and improve OOS risk and left-tail risk-adjusted returns
- Giudici et al. (2020): network models and Random Matrix Theory combined to improve robot advisory portfolio asset allocation
- Pichler et al. (2018): minimize systemic risk through network optimization

Other uses of clustering and networks analysis

- Vozlyublennaia and Wu (2018): measures mutual fund uniqueness using cluster analysis of fund returns
- Narabin and Boongasame (2018): cluster analysis of mutual funds data
- Lisi and Menardi (2015) and Lisi and Otranto (2010): clustering analysis to obtain classes of homogeneous funds with respect to risk levels
- Sakakibara et al. (2015): cluster mutual funds based on investment similarity instead of historical performances
- Forsberg et al. (2018): identifying hedge fund manager skill using cohorts (identified through clustering) that adjusts fund returns for return of cohort of funds applying same, or very similar, strategies.
- Razafitombo (2015): fund analysis and selection using distances and similarities between key performance measures based on clustering
- Zhang and Gencay (2019): mutual fund performance persistence linked to network-based analysis of how fund utilizes market information
- Eom and Park (2017): influence of factors affecting network connectivity is considerable on stocks with many links to other stocks in network

Other uses of clustering and networks analysis (cont.)

- Konstantinov and Rusev (2020): connections between global equity and bond funds from a network perspective; networks can amplify system-wide stress or inefficiencies in the factor bets
- Equity-bond allocations based on centrality scores, factor exposure, and hierarchical clustering of asymmetrically connected assets
- Konstantinov and Simonian (2020): analyze hedge fund market from perspective of a network of interacting individual funds
- Able to identify most important hedge fund styles to affect market and transmit systemwide shocks
- Baitinger and Papenbrock (2017b), Baitinger and Papenbrock (2017a): neglect of interconnectedness risk is not justified, because a) it has limited connection to conventional portfolio optimization input and b) investment strategies based on interconnectedness information outperform
- Balasubramanian et al. (2019): Market regime changes are connected to clustering of co-moving assets and breaking of already existing clusters

$\mathbb{ML}$ in $\mathbb{QWIM}$: forecasting of financial time series

Forecasting of time series

- Goal of econometric models is parameter estimation: produce good estimates of parameters β that underlie relationship between y and x
- Goal of (supervised) $\mathbb{ML}$ is to produce good predictions out-of-sample
- Mullainathan and Spiess (2017): "*Put succinctly, machine learning belongs in toolbox compartment marked $\widehat{y}$ rather than in the more familiar $\widehat{\beta}$ compartment*"
- Econometric models adjudicate explanatory power to specific variables; designed to decompose a combined effect as sum of individual effects
- Success of $\mathbb{ML}$ largely due to its ability to discover complex structure that was not specified in advance, and to fit complex and very flexible functional forms to data without simply overfitting.
- $\mathbb{ML}$ may enable new perspectives, such as recasting time series forecasting as an ordinal regression task Orozco et al. (2018)
- Comparing statistical and $\mathbb{ML}$ forecasting methods: Makridakis et al. (2018a), Makridakis et al. (2020), Januschowski et al. (2020), Benidis et al. (2020), Lim and Zohren (2020), Cerqueira et al. (2019), and Hyndman and Athanasopoulos (2020)

Forecasting of financial time series (cont.)

- Traditional methods for time series forecasting include univariate Autoregressive (AR), univariate Moving Average (MA), Autoregressive Integrated Moving Average (ARIMA): Hyndman and Athanasopoulos (2020)
- Siami-Namini and Namin (2018), Qian (2017): $\mathbb{ML}$ (in particular Deep Learning) may outperform ARIMA for financial time series forecasting
- Heaton et al. (2017): deep learning can detect and exploit data interactions invisible (at least currently) to any existing financial economic theory
- Siami-Namini and Namin (2018): Long Short-Term Memory (LSTM) delivers 85% reduction in error rates compared to ARIMA
- Bao et al. (2017): combination of wavelet transforms, stacked autoencoders and LSTM outperforms LTSM, WT-LTSM and recurrent neural network RNN
- Fischer and Krauss (2017): LSTM good out-of-sample prediction of directional movements for constituents of $S\&P$ 500
- Sezer et al. (2020), Alonso et al. (2020): financial time series forecasting with $\mathbb{DL}$

Forecasting of financial time series (cont.)

- Superiority of $\mathbb{ML}$ not apparent when it comes to typical time series forecasting, where data availability is often limited
- M4 competition: Makridakis et al. (2019) and Makridakis et al. (2018b)
  - many (100,000) and diverse time series, mostly medium-frequency data (daily), and some low-frequency data
  - requested both point forecasts and prediction intervals
- Performance for time series forecasting has to be assesed through a combination of metrics Botchkarev (2019) and Gasthaus et al. (2019)
- Januschowski et al. (2020): more practical to compare $\mathbb{ML}$ and statistical forecasting from other perspectives:
  - global versus local methods
  - probabilistic versus point forecasts
  - data driven versus model-driven
  - ensemble versus single models
  - explanatory/interpretable versus predictive

Combinations of forecasting methods

- Combinations of forecasting methods appear to have best results for forecasting of time series of similar granularity as the ones used in $\mathbb{QWIM}$: Qian et al. (2019), Gibbs (2017), and Mancuso and Werner (2013), Atiya (2020), Dantas and Oliveira (2018), Elliott (2017), and Hsiao and Wan (2014), Ma et al. (2018), Thomson et al. (2019) and Post et al. (2019), Shaub (2020)
- First place at M4 Competition Smyl (2020): combination of exponential smoothing ES and LTSM: ES for main components of individual series (seasonality, level), while LSTM for non-linear trends and cross-learning
- Second place at M4 Competition Montero-Manso et al. (2020): combination of 9 statistical and 1 $\mathbb{ML}$ methods, with optimal weights of combination obtained through $\mathbb{ML}$ (gradient boosting)
- Improving on winning entry Smyl (2020) for M4 competition quite challenging task; only a few specialized algorithms have succeeed
- Pure $\mathbb{ML}$ methods (e.g., N-BEATS, GluonTS) discussed in next slides
- More on combinations of $\mathbb{ML}$ and statistical: Joshi (2019), McDonald et al. (2014) and Karathanasopoulos et al. (2017), Gilliland (2020), Allende and Valle (2017), Zhao and Feng (2020)

Combinations of forecasting methods (cont.)

- Pike and Vazquez-Grande (2020): combination weights computed using ML trained on models' past forecast errors
- Bauer et al. (2020b), Bauer et al. (2020a): Telescope is a hybrid multi-step-ahead forecasting approach
  - based on time series decomposition forecasting
  - extracts intrinsic features based on STL decomposition + Fourier
  - XGBoost for composing the forecast component
  - can be used with or without a recommendation system
- Compared to hybrid forecasting methods (including Smyl (2020)), ML methods, established statistical methods, results show that Telescope has best forecast accuracy (based on sMAPE), lowest forecast error variability, much faster than any of the next 3 best (in terms of forecast accuracy) forecasting methods

Exploiting cross-series information

- Athanasopoulos et al. (2019) and Athanasopoulos and Kourentzes (2020): time series are often hierarchical and correlated
- Collection of time series can exhibit dependency relationships between individual time series that can be leveraged in forecasting
  - local co-variate relationships
  - indirect relationships through shared latent causes
  - subtle dependencies through smoothness, temporal dynamics, and noise characteristics
- Instead of one model per each time series, global model is developed exploiting information from many time series simultaneously
- Spiliotis et al. (2020) and Abolghasemi et al. (2019): $\mathbb{ML}$ applied to time series hierarchical forecasting may outperform traditional models
- Sirignano and Cont (2018): model trained on data from all stocks outperforms models trained on time series of any individual stock
- Impressively, results hold for stocks NOT included in training sample

Exploiting cross-series information (cont.)

- Hewamalage et al. (2019): survey of global models in context of deep neural networks.
- Smyl (2020): global model concept with local parameters
- Bandara et al. (2019): clustering groups of related time series, with global model developed per each cluster.
- Salinas et al. (2020): cross-series information applied to Probabilistic Forecasting with Autoregressive Recurrent Networks (DeepAR) model
- Oreshkin et al. (2020): Neural basis expansion analysis (N-BEATS) model
- Wang et al. (2019b): global Deep Neural Networks DNN backbone and local probabilistic graphical models for computational efficiency
- Rangapuram et al. (2018), Wang et al. (2019b): parameterize per-time-series linear state space model with a jointly-learned recurrent neural network

Results on M4 datasets

- Oreshkin et al. (2019): deep neural architecture based on backward and forward residual links and a very deep stack of fully-connected layers.
- N-BEATS: Neural basis expansion analysis.

Table 1: Performance on the M4 test set, sMAPE metric

	Yearly (23K)	Quaterly (24K)	Monthly (48K)	Others (5K)	Average (100K)
Best pure ML	14.397	11.031	13.973	4.566	12.894
Best statistical	13.366	10.155	13.002	4.682	11.986
Best ML/TS combination	13.528	9.733	12.639	4.118	11.720
DL/TS hybrid, M4 winner	13.176	9.679	12.126	4.014	11.374
NBEATS-G	12.855	9.378	12.130	3.979	11.229
NBEATS-I	12.823	9.418	12.048	4.199	11.203
NBEATS-I+G	12.812	9.372	12.064	4.063	11.190

Source: Oreshkin et al. (2019)

Results on M4 datasets (cont.)

Table 2: Performance on the M4 test set, OWA and M4 Rank

	Yearly (23K)	Quaterly (24K)	Monthly (48K)	Others (5K)	Average (100K)	Rank
Best pure ML	0.859	0.939	0.941	0.991	0.915	23
Best statistical	0.788	0.898	0.905	0.989	0.861	8
Best ML/TS combination	0.799	0.847	0.858	0.914	0.838	2
DL/TS hybrid, M4 winner	0.778	0.847	0.836	0.920	0.821	1
NBEATS-G	0.755	0.814	0.823	0.876	0.799	
NBEATS-I	0.753	0.8219	0.820	0.911	0.799	
NBEATS-I+G	0.752	0.814	0.819	0.889	0.797	

Source: Oreshkin et al. (2019)

Table 3: Comparison betwen DeepAR and Smyl method for M4 dataset

Metric	DeepAR	Smyl
sMAPE	0.1192	0.1137
MASE	1.500	1.54
OWA	0.837	0.821
MSIS	12.07	12.23

Source: Salinas et al. (2019)

Results on M4 datasets (cont.)

- Loning et al. (2019): sktime, a general $\mathbb{ML}$ framework for time series, with
  - scikit-learn compatible interfaces
  - tools to build, tune and evaluate composite models
- Loning and Kiraly (2020): framework enhanced to both replicate and extend key results from M4 forecasting study
- Notable results:
  - perfect replication for all naive forecastors
  - small but often statistically significant differences for statistical models
  - large and statistically significant differences for MLP and RNN (likely due to improvements in sk-time and TensorFlow)
  - simple hybrid approaches (such as residual boosting) can boost the performance of statistical models

Empirical asset pricing

- Karolyi and Van Nieuwerburgh (2020), Hou et al. (2021), Nietert and Otto (2020), Barras (2019), Fama and French (2020), Fletcher (2019b), Bektic et al. (2020), Lettau and Pelger (2020), Chen and Zimmermann (2020), Maio (2019), Zaffaroni (2019), Chib et al. (2020): current state-of-the-art on empirical asset premia
- Harvey et al. (2016), Hwang and Rubesam (2019), Feng et al. (2020), Harvey and Liu (2020a), Pukthuanthong et al. (2019): asset pricing literature has produced hundreds of potential risk factors
- Chib and Zeng (2020), Harvey and Liu (2020c), Fletcher (2019a), Messmer and Audrino (2020), Green et al. (2017): Which characteristics really provide independent information?
- Literature mainly builds on linear or linear-like models, due to simplicity, transparency and computational efficiency
- Collot and Hemauer (2021), Giglio and Xiu (2019): Concerns for risk premia estimation: potential omission of factors, measurement error, nonlinearity, clustering or dependency structure

Empirical asset pricing (cont.)

- Suhonen et al. (2017), Shi and Li (2021), Xu (2020), Bryzgalova et al. (2020), Gospodinov and Robotti (2021), Freyberger et al. (2020): current approaches still have challenges
- $\mathbb{ML}$ relatively well-equipped to address these concerns, since no prior imposed on functional relation describing expected return process, but "learn" this non-linear dependence structure.
- Gu et al. (2020): $\mathbb{ML}$ improved performance compared to traditional methods for both cross section and time series stock return prediction
- Predictive gains traced to inclusion of nonlinear interactions, while $\mathbb{ML}$ methods agree on a fairly small set of dominant predictive signals
- Gu et al. (2021): autoencoder asset pricing model delivers OOS errors far smaller compared to other leading factor models
- Bianchi et al. (2021): $\mathbb{ML}$ delivers useful bond risk premia forecasts
- Dixon and Polson (2020): $\mathbb{DL}$ fundamental factor models capture non-linearity, interaction effects and non-parametric shocks

Empirical asset pricing (cont.)

- Feng et al. (2018): extraction of nonlinear factors information through $\mathbb{DL}$, with statistical factor models special cases of shallow learners
- Coqueret (2020a): importance of memory in $\mathbb{ML}$-based models relying on firm characteristics for asset pricing
- Drobetz and Otto (2020): evaluates performance of $\mathbb{ML}$-based forecasting stock returns
- Rapach and Zhou (2020): $\mathbb{ML}$ forecasts using time series and cross-sections of stock returns
- Coqueret and Guida (2020): $\mathbb{ML}$-based models for factor investing
- Andreini et al. (2020): $\mathbb{ML}$ used to encode information available from hundreds of macroeconomic and financial time-series into a handful of unobserved latent states, which can still be interpreted as in a standard factor models

Empirical asset pricing (cont.)

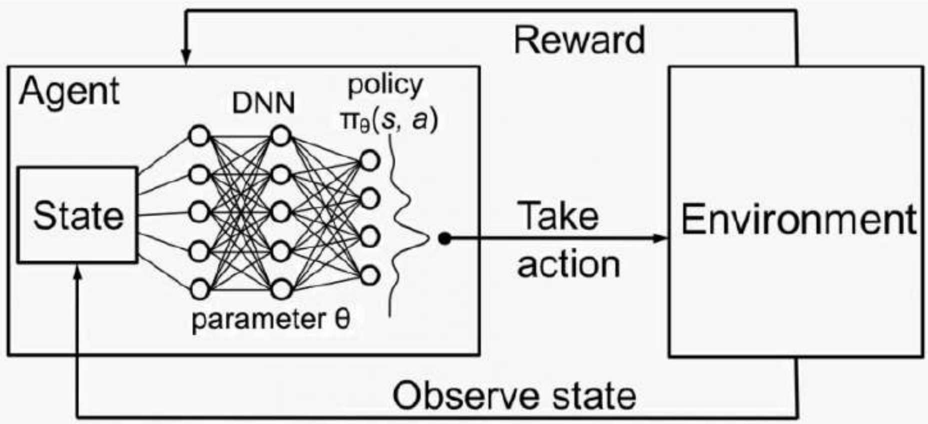
- Grammig et al. (2020): $\mathbb{ML}$ able to predict returns and generate profits in 33 international markets, while using fewer variables as inputs compared to previous studies
- Jurczenko et al. (2020), Guida (2019), Lopez de Prado (2020a): advances and promising results when using $\mathbb{ML}$ for empirical asset pricing
- However, while $\mathbb{ML}$-based risk premia has many advantages, it also has practical challenges
- Grammig et al. (2020): in a low signal-to-noise environment theory-based risk premia might deliver better results compared to $\mathbb{ML}$-based risk premia
- Leung et al. (2020): Despite statistical advantage of $\mathbb{ML}$ model predictions, corresponding economic gains somewhat limited, largely dependent on ability to take risk and implement trades efficiently.

$\mathbb{ML}$ in $\mathbb{QWIM}$: reinforcement learning $\mathbb{RL}$

Reinforcement learning $\mathbb{RL}$

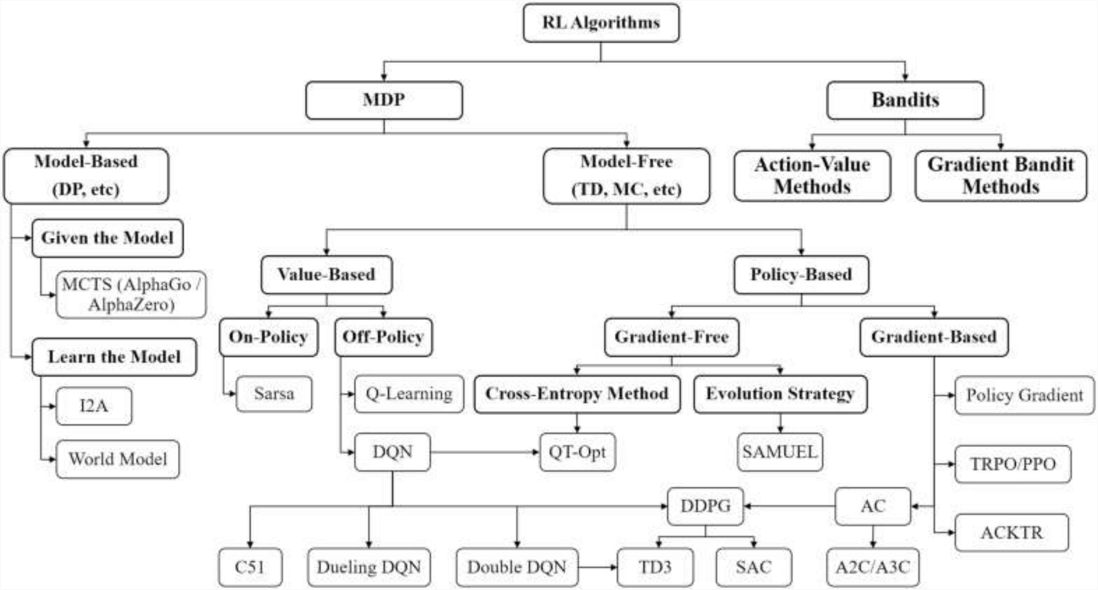
- $\mathbb{RL}$ "learns" how to attain a complex objective (goal), with agent trained in environment by performing actions and adapting to results
- Interactive process, as agent actions actively change its environment
- Deep $\mathbb{RL}$: agents construct and learn their own knowledge through deep learning of neural networks.

Figure 1: (Deep) Reinforcement learning diagram



Source: Campos (2018)

Taxonomy of $\mathbb{RL}$ algorithms



Source: Zhang and Yu (2020)

Financial applications of $\mathbb{RL}$

- $\mathbb{RL}$ can solve financial applications of intertemporal choice, such as:
  - pricing and hedging of financial derivatives
  - investment and portfolio allocation
  - buying and selling securities subject to transaction costs
  - optimal dynamic trading strategy
  - optimization of tax consequences
- There are already good results for $\mathbb{RL}$ and $\mathbb{DRL}$ in trading and hedging: Kolm and Ritter (2019), Ritter (2019), Kolm and Ritter (2020b), Dixon et al. (2020)
- Buehler et al. (2019): $\mathbb{DRL}$ framework for hedging a portfolio of derivatives in presence of transaction costs, liquidity constraints or risk limits
- Du et al. (2020): $\mathbb{DRL}$ for option replication subject to discrete trading, round lotting, and nonlinear transaction costs
- Cao et al. (2020): $\mathbb{DRL}$ for optimal hedging under transaction costs
- Carbonneau (2020), Carbonneau and Godin (2020), Cao et al. (2021), Benhamou et al. (2020): $\mathbb{DRL}$ for hedging of financial derivatives (including long term)

Using $\mathbb{RL}$ in $\mathbb{QWIM}$

- Fischer (2018): $\mathbb{RL}$ allows to combine “prediction" and “portfolio construction" tasks into one integrated step, providing better alignment to investor objectives
- Noguer i Alonso and Srivastava (2020): $\mathbb{DRL}$ for model-free asset allocation
- Sato (2019): model-free $\mathbb{RL}$ for financial portfolios
- Irlam (2020): $\mathbb{RL}$ is capable of handling real-world complexity of financial planning; taxes, inflation, longevity, etc.
- Irlam (2020): financial and retirement planning via $\mathbb{DRL}$ outperforms
- Benhamou et al. (2021): bridge gap between $\mathbb{DRL}$ and MPT; $\mathbb{DRL}$ maps directly market conditions to actions by design
- Benhamou et al. (2021): augmented asset management via $\mathbb{DRL}$
- Hu and Lin (2019), Cong et al. (2020): $\mathbb{DRL}$-based portfolio management
- Alsabah et al. (2021): using $\mathbb{RL}$, roboadvisor “learns” over time investor risk preference by observing her portfolio choices under different markets.

Using $\mathbb{RL}$ in $\mathbb{QWIM}$: Goals-based investing

- Goal Based Investing $\mathbb{GBI}$: investment approach with performance measured by probabilities of success in achieving investor financial goals (at given time horizons and for specified priority levels)
- Objective is to invest systematically and to manage a portfolio consistent with investor's risk profile and time horizon of goals.
- References for $\mathbb{GBI}$: Kim et al. (2020), Roncalli (2019), Wang et al. (2011), Deguest et al. (2015), Janssen et al. (2013), Dempster et al. (2016), Shalett (2015), Brunel (2011), Parker (2016b), Parker (2016a), Chhabra (2015), Parker (2020)
- Grealish and Kolm (2021), D'Acunto and Rossi (2021): roboadvisors aim to become more holistic; need to incorporate $\mathbb{GBI}$, $\mathbb{ML}$, and more
- Combination of $\mathbb{RL}$ and $\mathbb{GBI}$ very promising, since investor goals in $\mathbb{GBI}$ can be connected to reward function in $\mathbb{RL}$
- $\mathbb{GBI}$ portfolio optimization, viewed as optimal control problem performed within data-driven world, can be solved via $\mathbb{RL}$
- Depending on desired purpose, $\mathbb{RL}$-based solutions for $\mathbb{GBI}$ may need to incorporate techniques from $\mathbb{RL}$, $\mathbb{DRL}$, $\mathbb{IRL}$

Using $\mathbb{RL}$ in $\mathbb{QWIM}$: Goals-based investing (cont.)

- Das and Varma (2020): use $\mathbb{RL}$ to solve for $\mathbb{GBI}$, with the solution converging to the solution obtained using dynamic programming
- Dixon and Halperin (2020): $\mathbb{GBI}$ using G-Learner, which
  - operates with explicitly defined one-step rewards
  - does not assume a data generation process
  - is suitable for noisy data
- Dixon and Halperin (2020): GIRL extends G-Learner within context of Inverse Reinforcement Learning $\mathbb{IRL}$, where rewards collected by agent are not observed, and should instead be inferred
- Given current state-of-the-art for $\mathbb{RL}$, incorporating multiple goals with their corresponding priority levels within a $\mathbb{RL}$-based solution for $\mathbb{GBI}$ appears to be a significant modeling challenge
- It is likely that such a solution will leverage hierarchical $\mathbb{RL}$ and curiosity-driven exploration: Zhao et al. (2019), Rafati and Noelle (2019), Xing (2019), Colas et al. (2019), Roder et al. (2020).

$\mathbb{ML}$ in $\mathbb{QWIM}$: importance of market states and regimes

Market states in QWIM

- There are two separate market states Peters (2015), Munnix et al. (2012), Pharasi et al. (2020) Ladekarl et al. (2019), Zakamulin (2020), Zaremba et al. (2020):
  - low uncertainty (relatively stable and resilient) market
  - high uncertainty (relatively chaotic and fragile) market
- Markets in **LOW UNCERTAINTY** state:
  - statistically well behaved
  - can be modeled using standard statistical tools
  - volatility stable and low, correlations relatively stable
  - tail events ($\geq$3 std deviations in either direction) quite rare.
- Markets in **HIGH UNCERTAINTY** state:
  - not statistically well behaved
  - vols and correlations change significantly on regular basis
  - Tail events happen with much more regularity
- To account for the two market states, practitioners use a relatively similar concept of **"risk on, risk off"**: Lee (2012), Ma (2015), Dapena et al. (2019)

Structural breaks: market regimes

- The **HIGH UNCERTAINTY** state can incorporate multiple instances and types of significant changes in time series: market regimes, changepoints, bubbles/crashes
- Regime changes, some transitory, some recurring (recessions versus expansions) some permanent (structural breaks), are prevalent across a wide range of financial markets and in behavior of many macro variables Ang and Timmermann (2012), Hamilton (2016), and Sheikh and Sun (2012)
- Examples of regimes considered in academia and/or practitioners:
  - bull vs. bear market regimes
  - inflationary vs. recessionary regimes
  - high vs. low volatility regimes
  - mean reverting vs. trending regimes
- Kritzman et al. (2012) and Blitz and Van Vliet (2009): Regime shifts are challenging for investors because they cause portfolio performance, risk and behavior to depart significantly from ranges implied by long-term averages of means and covariances

Structural breaks: Market regimes (cont.)

- Regime-based asset allocation was shown to deliver improved performance and risk profile Masih (2012), Vo and Maurer (2013), SSgA Research (2015), Nystrup et al. (2018b), Briere and Signori (2012), Nystrup et al. (2015), Kritzman et al. (2012), Kaya (2017), Teiletche et al. (2017), Sheikh and Sun (2012), Jurczenko and Teiletche (2018), Lewin and Campani (2020), Lezmi et al. (2018), Fons et al. (2021), Collin-Dufresne et al. (2020),
- Kinlaw and Turkington (2014): periods characterized by correlation surprise lead to higher risk and lower returns to risk premia than periods characterized by typical correlations.
- Ammann and Verhofen (2006): empirical analysis indicates two clearly separable regimes with different mean returns, volatilities and correlations
- Papenbrock and Schwendner (2015): distinct correlation regimes detected
- Bernhart et al. (2011): correlations between (and within) asset classes are far from being stable, and they vary significantly
- Papenbrock and Schwendner (2015) and Simonian and Wu (2019b): regime-based investing, based on spectral clustering

Forecasting performance depends on market regimes

- Asset allocation is critically dependent on out-of-sample risk premium.
- Question: *Is superior econometric predictability across the business cycle synonymous with predictability at all times*?
- Answer is important since there is asymmetric behavior:
  - **gains during good periods** (when economic benefits may be reduced for high risk-averse and leverage-constrained investors)
  - **losses during bad periods** (when it matters the most for asset allocators to retain assets and their client base intact).
- Li and Piqueira (2019), Dal Pra et al. (2018), Salhi et al. (2016), Nalewaik (2015), Pereiro and Gonzalez-Rozada (2015), Chincoli and Guidolin (2017) Huang et al. (2017),Odendahl et al. (2020), Tsiakas et al. (2020), Hammerschmid and Lohre (2018) : predictability of returns (and of other investment variables) is regime-dependent.
- Baltas and Karyampas (2018): recent advanced forecasting models more accurate in up markets, during expansions, and during low-volatility periods; not necessarily in down markets, during recessions, and during high-volatility periods.

Forecasting performance depends on market regimes (cont.)

- Perfomance of investment strategies greatly enhanced with introduction of regime switching models $\mathbb{RSM}$s Costa and Kwon (2020), Nystrup et al. (2015), Nystrup et al. (2018b), Nystrup et al. (2018a), Flint and Mare (2019), Gokhale et al. (2019), Alberico et al. (2018), Collin-Dufresne et al. (2020)
- $\mathbb{RSM}$s characterize market states using estimates of parameters of some underlying model, and use a transition matrix to quantify probability of moving from one state to another.
- $\mathbb{ML}$ may be effective at detecting change (even in chaotic systems Wolchover (2018)), for example through robust anomaly detection Bulusu et al. (2020), Braei and Wagner (2020)
- It can ban be enhanced to compute probability of observation in previously observed "market regimes" (defined as clusters in $\mathbb{ML}$)
- Thus clustering algorithms can identify regimes in datasets
- What they have in common with regular regime switching models is ability of producing probabilities of "switching" into another regime
- $\mathbb{ML}$ can also detect preconditions of a break

Identifying market regimes

- Unsupervised learning algorithms can be used for the determination of "natural" clusters and the grouping of market data according to predefined similarity criteria.
- Supervised learning algorithms can be "trained" with known data sets to classify market situations. Then, new data – and especially the current situation – can be assigned to the market phases.
- Mulvey and Liu (2016): identify market regimes via trend filtering
- Chazal and Michel (2017): Topological Data Analysis TDA used to automatically cluster thousands of variables spanning market, macroeconomic, and sentiment data and to create create visual, topological summaries that reveal the key characteristics of the current regime as well as the similarities to past regimes.
- Gidea and Katz (2018): use TDA (persistence homology) to detect and quantify topological patterns that persist across multiple scales; metrics of persistence landscapes exhibit strong growth prior to the primary peak, which ascends during a crash.

Structural breaks: bubbles and crashes

- Alves et al. (2018): Chaotic systems of the real world are comparable to stock market indices evolution.
- Sornette and Cauwels (2015): financial bubble defined as a period of unsustainable growth, when the price of an asset increases ever more quickly, in a series of accelerating phases of corrections and rebounds.
- Filimonov and Sornette (2013) and Sornette (2014): log-periodic power law singularity (LPPLS) model identifies well bubbles and crashes
- Demos and Sornette (2017): LPPLS provides overwhelming evidence that the beginning of bubbles is much better constrained that their end.
- Demirer et al. (2019): LPPLS framework successfully captures, ex-ante, most prominent bubbles across different time scales (Black Monday, Dot-com, and Subprime Crisis).
- Change point detection (CPD) is the problem of finding abrupt changes in data when a property of the time series changes
- Segmentation, edge detection, event detection, and anomaly detection are similar concepts within $\mathbb{ML}$ space.

Structural breaks: changepoints

- Ruff et al. (2020), Blazquez-Garcia et al. (2020), Foorthuis (2020), Braei and Wagner (2020), van den Burg and Williams (2020), Aminikhanghahi and Cook (2017): surveys of algorithms for change point detections
- Wu and Keogh (2020), Goswami et al. (2018): existing transition detection methods do not rigorously account for time series uncertainties
- Lee et al. (2018): traditional changepoint detection methods look for statistically-detectable boundaries (defined as abrupt variations in generative parameters of a data sequence)
- However, breakpoints may occur on boundaries non-trivial to detect with these statistical methods, but detectable using $\mathbb{DL}$
- Chalapathy and Chawla (2019), Chalapathy et al. (2020), Wang et al. (2020a),Lee et al. (2018): $\mathbb{DL}$ for anomaly detection
- Bashar and Nayak (2020), Geiger et al. (2020): Generative Adversarial Networks for anomaly detection

$\mathbb{ML}$ in $\mathbb{QWIM}$: other practical applications

Generating synthetic data for $\mathbb{QWIM}$

- Significant need for additional data in $\mathbb{QWIM}$
  - for adequately training $\mathbb{ML}$ models
  - to tackle unbalanced datasets
  - for constructing a very comprehensive set of scenarios incorporating most scenarios which were not observed, yet are plausible and consistent with market and investor behavior and with economic intuition
  - for data anonymization and preserving of data privacy
  - for comprehensive testing of investment strategies and portfolios
  - for constructing much more robust portfolios using scenario-based optimization
  - for better risk management of investment portfolios
- Assefa et al. (2019): synthetic data defined as data obtained from a generative process that learns properties of real data
- Solutions through $\mathbb{ML}$-based data augmentation **DAug**

Data augmentation and scenario generation

- Coulombe (2018): successful use in NLP of same **DAug** techniques already showing great success in artificial vision
- Fawaz et al. (2018a): **DAug** based on Dynamic Time Warping distance
  - drastically increase of accuracy of deep Convolutional Neural Networks on some time series $\mathbb{TS}$ datasets
  - significantly improvement of deep model's accuracy when used in an ensemble approach
- Kuchnik and Smith (2018): subsampling policies based on model influence and loss achieve a 90% reduction in augmentation set size while maintaining the accuracy gains of standard data augmentation
- Zhang et al. (2018c): DADA: Deep Adversarial Data Augmentation for Extremely Low Data Regime Classification
- Hernandez-Garcia and Konig (2018) and Da Silva and Shi (2019): generative model for financial $\mathbb{TS}$ able to create realistic paths that embed underlying structure of markets in a way stochastic processes could not

Generating synthetic financial time series

- Synthetic financial TS data is expected to have universal features, commonly referred to as stylized facts (Kristoufek (2018), Jilla et al. (2017), Cont (2001), Allen and Satchell (2014)), together with specialized features of a specific asset class or investment vehicle
- Buehler et al. (2020a): overview of currently used generative modelling approaches and performance evaluation metrics for financial TS
- For classical models, these stylized facts are often formulated in terms of distribution of returns
- Buehler et al. (2020a): This returns-based viewpoint (though still interesting) may not convey a sufficiently full picture for distributions of synthetic market paths obtained using generative modeling.
- Most promising ML methods to generate synthetic financial TS:
  - generative adversarial networks GANS
  - Variational Autoencoders VAEs
  - Restricted Boltzmann Machine RBM

Generating synthetic financial time series (cont.)

- Wiese et al. (2020): realistic multivariate financial $\mathbb{TS}$ based on GANs, able to capture longer ranging dependencies
- Buehler et al. (2020a): model based on VAEs works reliably even cases where amount of available training data is notoriously small
- Kondratyev and Schwarz (2019) and Kondratyev et al. (2020): RBM delivers synthetic market data with similar probability distribution of original data, controlling autocorrelation and non-stationarity in generated $\mathbb{TS}$
- Takahashi et al. (2019): GANs produce $\mathbb{TS}$ with desired properties: linear unpredictability, heavy-tailed price return distribution, volatility clustering, leverage effects, coarse-fine vol correlation
- Koshiyama et al. (2019): GANs for financial trading strategies
- Mariani et al. (2019): PAGAN: Portfolio Analysis with GANs
- Marti (2020): GANs to generate realistic financial correlation matrices
- De Meer Pardo and Lopez (2020): training $\mathbb{ML}$ on synthetic data generated by Wasserstein GAN with gradient penalty mitigates overfitting

Generating synthetic financial time series (cont.)

- Franco-Pedroso et al. (2019a): virtual multivariate scenarios of financial data of arbitrary length and asset composition, retaining volatility clustering, cross-asset relationships, and their changes over time
- Franco-Pedroso et al. (2019b): While posterior distribution of scenarios has to be consistent with that of real data, it is not sufficient because other properties related to evolution of time series (e.g., autocorrelation of returns) not being evaluated or unknown as being discriminant
- Franco-Pedroso et al. (2019b): evaluation framework to
    - quantify degree of realism of simulated financial time series, whatever simulation method might be
    - discover and improve unknown characteristics not being properly reproduced by analyzed methods
- Lezmi et al. (2020): Boltzmann machines and GANs to generate synthetic data preserving first four statistical moments, stochastic dependence between different dimensions (copula structure) and across time (autocorrelation).

Portfolio construction

- As discussed, network analysis and clustering could be leveraged to construct a more robust and diversified portfolio
- Snow (2020): $\mathbb{ML}$ can help with most portfolio construction tasks, such as idea generation, alpha factor design, asset allocation, weight optimization, position sizing, as well as with testing
- Snow (2020): identifies different weight optimization methods within
  - supervised learning framework
    - traditional linear approach using OLS, Ridge, LASSO
    - nonlinear deep learning approach using autoencoders
    - Bayesian sentiment method
  - unsupervised learning framework
    - principal component analysis
    - hierarchical clustering analysis
    - network graphs
  - reinforcement learning framework
    - Deep Determinist Policy Gradient

Testing investment strategies and portfolios

- Harvey et al. (2020b), Giglio et al. (2021), Chordia et al. (2020), Vincent et al. (2020), Perumal and Flint (2018): comparison of various methodologies to test investment strategies
- Backtesting is at high risk of spurious accuracy due to data-mining bias present from considering multiple rules concurrently over same history.
- Harvey and Liu (2018), Barras et al. (2010), Barras et al. (2019), Bailey et al. (2017), Bailey et al. (2014), Bryzgalova et al. (2020), Harvey and Liu (2020b), Hens et al. (2020), Dichtl et al. (2020): majority of investment strategies subject to backtest overfitting (outperforms during backtest, underperforms in practice)
- When strategy selection is done under multiple testing (running many alternative configurations), backtest likely false discovery.
- $\mathbb{ML}$ can be trained to scan billions of data signals in order to design millions of different virtual investment strategies
- Lopez de Prado (2018): Due to backtest overfitting, most quantitative investment firms invest in false discoveries

Testing investment strategies and portfolios (cont.)

- Lopez de Prado (2020a), Lopez de Prado (2018):These false discoveries may have been prevented if academic journals and investors demanded that any reported investment performance incorporates the false positive probability, adjusted for selection bias under multiple testing
- Fabozzi and Lopez de Prado (2018): propose a template to fairly disclose multiple tests involved in testing investment strategies
- Chordia et al. (2020): apply multiple hypothesis testing techniques that account for cross-correlations in signals and returns to produce t-statistic thresholds that control proportion of false discoveries.
- very small number of strategies survive when combining statistical criteria with economic considerations
- Aparicio and Lopez de Prado (2018): model confidence set MCS (see Hansen et al. (2011)) may not be applicable to investment strategies testing since:
    - it is not robust to multiple testing
    - it requires a very high signal-to-noise ratio to be useful.

Testing investment strategies and portfolios (cont.)

- Kellner and Rosch (2020): Bayesian re-interpretation of "significant" empirical financial research indicate rather weak evidence for inferences made under frequentistic framework
- Testing needs to go beyond simple historical data and regular bootstrap
- Harvey and Liu (2015): enhanced bootstrap for multiple comparisons, need to incorporate fictitious yet plausible market scenarios
- Rebonato (2017) describe engineering challenges for testing framework, to incorporate both '***scenario analysis***' and '***stress testing***'.
- $\mathbb{ML}$ can generate such additional scenarios (see Data Augmentation)
- Denev and Angelini (2016), Denev et al. (2018): Probabilistic Graphical Models can capture interconnectedness within a stress testing context
- Gao et al. (2018): combination of probabilistic causation and $\mathbb{ML}$ can simulate stress testing scenarios with higher accuracy and lower computational complexity than Monte Carlo simulations.

Testing investment strategies and portfolios (cont.)

- Arnott et al. (2019): Backtesting Protocol in the Era of Machine Learning
  - avoid HARKing (Hypothesizing After Results are Known)
  - be aware of multiple testing problem
  - be aware of data integrity
    - outlier exclusions and data transformations set in advance
    - results robust to minor changes in transformations
  - be honest with cross-validation
    - avoid modifying in-sample model to fit OOS data
    - ensure OOS analysis is representative of live trading
    - ensure realistic market frictions such as transactions costs
  - awareness of model dynamics
    - model has to be robust to structural changes
    - minimize model overfitting and tweaking based on live results
    - minimize practical risks such as overcrowding
  - avoid complexity
    - use simplest yet practicable model
    - be able to explain results
  - need to have a scientific research culture
    - reward good proceses, not good results
    - are results inconsistent with other research?

Factor-based investment strategies

- Traditional (linear) risk factor models may miss (partially or entirely) key features due to model misspecification
- $\mathbb{ML}$ very good at identifying patterns in a high dimensional space
- These patterns asssociate features (factors) with outcomes
- While nature of relationships can be quite complex, primary focus of analysis can be geared towards finding most important features
- $\mathbb{ML}$ will always find patterns, even if none or minimal in reality
- Likelihood of overfitting is high, due to small ratio of size of samples (returns) relative to the number of potential factors.
- To avoid "data mining" (spurious relationships that fail to generalize out of sample) $\mathbb{ML}$ has to be paired with economic intuition and financial knowledge

Factor-based investment strategies (cont.)

- Typically two types of methods to identify return predictors:
  - (conditional) portfolio sorts based on one or multiple characteristics, such as size or book-to-market
  - linear regression (panel data or in spirit of Fama-MacBeth)
- $\mathbb{ML}$ found to significantly improve selection process, since
  - Portfolio sorts subject to curse of dimensionality
  - linear regressions, as strong functional-form assumptions, are sensitive to outliers
  - $\mathbb{ML}$ accommodates many more potential predictor variables
  - $\mathbb{ML}$ accommodates richer specifications (e.g. nonlinear)
- Need to identify most relevant features (known as risk factors) and to analyze whether a new factor adds explanatory power
- Feature selection FS is an essential component for $\mathbb{ML}$

Factor-based investment strategies (cont.)

- FS methods aim at ranking and selecting a subset of relevant features according to their degrees of relevance, preference, or importance.
- Feng et al. (2020), Chernozhukov et al. (2015): conventional $\mathbb{ML}$ variable selection techniques need to be augmented to explicitly account for model-selection mistakes (like omitted variables)
- Feng et al. (2020): procedure selects factors that are useful in
  - explaining the cross section of expected returns
  - mitigating the omitted variable bias problem
- Freyberger et al. (2020): use group LASSO to
  - find which characteristics have incremental predictive power for expected returns, given the other characteristics.
  - estimate effect of important characteristics on expected returns without imposing a strong functional form.

Factor-based investment strategies (cont.)

Table 4: Output of selection procedures based on augmented $\mathbb{ML}$

Results	Article
14 out of 99	Feng et al. (2017)
13 out of 62	Freyberger et al. (2018)
14 out of 68	Messmer and Audrino (2017)

- Feng et al. (2018): consider empirical asset pricing problem from perspective of finding alphas in large panel datasets of return with a view to measuring investment performance, as described in Harvey and Liu (2019), Harvey and Liu (2018)
- From this perspective, empirical asset pricing can be recasted into a different solution: searching for a dimension-reduction model from thousands of securities, which is then done through deep learning
- With better measurement through machine learning, risk premia becomes less shrouded in estimation error

Factor-based investment strategies (cont.)

- Factor-based investment strategies particularly suited for $\mathbb{ML}$, because:
  - measurement of asset's risk premium is in fact prediction
  - large number of candidate conditioning variables for risk premium
  - nonlinearity is likely necessary
  - analysis is required to determine most appropriate functional form of risk factor model
- Gu et al. (2020): comparative analysis of $\mathbb{ML}$ for measuring asset risk premia
- Freyberger et al. (2020): nonlinearities are important; with same set of 11 factors selected by nonlinear model, $\mathbb{ML}$ has out-of-sample Sharpe ratio SR three times larger compared to a linear model
- Feng et al. (2018): predictability of returns explained by nonlinear factors
- Fan et al. (2017): connection between sufficient factor forecasting and deep learning

Additional applications of $\mathbb{ML}$ in $\mathbb{QWIM}$

- Kakushadze and Yu (2020): $\mathbb{ML}$ to extract factors underlying Treasury yields
- Kakushadze and Yu (2020): LSTM and TCN outperforms analysts in prediction of future earnings per share (EPS)
- Hou et al. (2020): LSTM-DNN hybrid model to integrate company fundamentals information into time-series forecasting
- Gotze et al. (2020): $\mathbb{ML}$ improves catastrophe CAT bond pricing models
- Bolhuis and Rayner (2020): nowcast/forecast macroeconomic variables
- Abouseir et al. (2020): $\mathbb{ML}$ allows to better integrate macroeconomic data into multi-asset allocation
- Jaeger et al. (2020): use "explainable AI" to compare robustness of diversification strategies in asset allocation, and to extract explanations
- Koshiyama et al. (2020): review of $\mathbb{AI}$ and $\mathbb{ML}$, their computational strengths and weaknesses, and their future impact on Capital Markets
- Downey (2020): ensemble $\mathbb{ML}$ to predict company fundamentals: earnings, free cash flow, EBITDA, Net Operating Profit After Taxes.

Additional applications of $\mathbb{ML}$ in $\mathbb{QWIM}$ (cont.)

- Nunes et al. (2020): bond yield forecasting using LSTM-LagLasso
- Li et al. (2020b): practitioner framework for implementation and interpretation of $\mathbb{ML}$ model predictions applied to investment portfolios.
- Model predictions are decomposed into linear, nonlinear, and interaction components, and evaluated using these components.
- It is shown that $\mathbb{ML}$ models reliably identify known effects and find new nonlinear relationships and interactions.
- Vrontos et al. (2021), Yazdani (2020): $\mathbb{ML}$ prediction of recessions
- Li and Zhang (2020): unified model for detecting, classifying and summarizing financial events
- Wolff and Echterling (2020), Borghi and Rossi (2020): $\mathbb{ML}$ for stock selection and for constructing a multifactor alpha model

Additional applications of $\mathbb{ML}$ in $\mathbb{QWIM}$ (cont.)

- Boudt et al. (2020): $\mathbb{ML}$ applied to tactical asset allocation
- Abouseir et al. (2020): integration of macroeconomic data into multi-asset allocation using $\mathbb{ML}$
- Bolhuis and Rayner (2020): $\mathbb{ML}$ to nowcast (and forecast) (macro)economic variables
- Li and Zhang (2020): unified model for detecting, classifying and summarizing financial events
- Nunes et al. (2020): use learned representations from $\mathbb{DL}$ networks to augment covariance estimation and to create tailored representations best suited to meeting varying financial objectives
- Zhou (2020): use graphical models for portfolio allocation
- Kilburn (2019a), Fornaro and Luomaranta (2020), Monokroussos and Zhao (2020), Babii et al. (2020): $\mathbb{ML}$ for nowcasting

Lessons from COVID-19 pandemic

- For many reasons, 2020 is an exceptional year.
  - COVID-19 virus has shown how vulnerable world is to a pandemic
  - "whatever-it-takes" response from central banks fastest ever seen
  - reminder that the highly unlikely is not the same as impossible, and the highly likely not same as certain
- Pinson and Makridakis (2021), Ioannidis et al. (2020), Taleb et al. (2021): significant challenges remain to deliver "good enough" forecasting methodologies
- Procacci et al. (2020), Cheng (2020): US market structure has dramatically changed during COVID outbreak, while volatility markets underreacted
- Harvey et al. (2020a): strategic risk management useful in market selloff
- Billio and Varotto (2020): economic impact and likely market consequences
- Cepoi (2020): asymmetric dependence between stock market returns and news during COVID-19 financial turmoil
- Lyócsa et al. (2020), Lyócsa and Molnár (2020), Salisu and Akanni (2020), Liu et al. (2020): role of fear and uncertainty in financial markets

Lessons from COVID-19 pandemic (cont.)

- Lipton and Lopez de Prado (2020b), Lipton and Lopez de Prado (2020a): convincing case for more nowcasting, less backtesting (or rather enhanced testing), and strategies that adapt to new regimes
- Nowcasting is becoming much more feasible due to combination of vastly larger amount of available data with $\mathbb{ML}$ advances
- Their list of recommendations includes
  - incorporate short-term 'nowcasting'
  - backtesting part of validation process, not of research process
  - need for analysis and methods (such as causal $\mathbb{ML}$, discussed later) towards a stronger link between cause and effect
  - rather than using "all-weather" investment strategies, identify investment strategies that perform well under defined market regimes
- Regime-based investing was already discussed
- Lopez de Prado (2020c): tactical investment algorithms TIAs

Lessons from COVID-19 pandemic (cont.)

- Lopez de Prado (2020c) described a process to construct TIAs
  - define each regime using a particular data-generating process
  - nowcast the probability that current observations are being drawn from each process
  - use those probabilities to build an ensemble portfolio of those optimal strategies
- As these probabilities shift from one data-generating process to another over time, the ensemble portfolio is dynamically adjusted and adapts to prevailing market conditions.
- Lipton and Lopez de Prado (2020b): During coronavirus sell-off, ensemble portfolio could have reduced allocations to models optimised for economic expansions and increased allocations to models optimised for economic recessions and market turmoil.

$\mathbb{ML}$ in $\mathbb{QWIM}$: practical challenges

Main challenges

- Challenges due to data Assefa et al. (2019), Buehler et al. (2020a)
  - insufficient data and/or somewhat poor data quality
  - need to satisfy privacy, fairness and regulatory requirements
- Financial datasets have distinctive stylized features Cont et al. (1997), Cont (2001), Kristoufek (2018), Jilla et al. (2017)
- Model overfitting and data mining risks higher for $\mathbb{ML}$ (when used inappropriately)
- Standard $\mathbb{ML}$ may not be “good enough” for financial time series
- Need to incorporate causality into $\mathbb{ML}$
- Stakeholders and regulators require significant levels of interpretability and/or explainability for $\mathbb{ML}$
- Tuning $\mathbb{ML}$ hyperparameters needs to be robust with respect to stylized features of financial time series (such as market regimes)
- Training and testing datasets needs to be representative of realistic investment and trading conditions (including stress periods)

Challenge: insufficient data

- Not having sufficient data (due to medium to low sampling frequency and/or limited historical data) will have an impact on
  - adequately training $\mathbb{ML}$ models
  - tackling unbalanced datasets
  - data anonymization and preserving of data privacy
  - comprehensive testing of investment strategies and portfolios
  - robust portfolio construction
  - portfolio risk management
- Synthetic data can be obtained through
  - $\mathbb{ML}$-based data augmentation DAug
  - data similarity with data in other time periods
  - combine observations with expert views
  - scenario generation
  - construction of stress scenarios
  - agent based modeling

Methods to generate synthetic data

- Methods to deliver consistent and realistic synthetic data include
  - $\mathbb{ML}$-based data augmentation DAug (discussed in previous section)
  - time similarity with data in other periods: Marrs (2019), Bai et al. (2019), Echihabi et al. (2018), Kang et al. (2019), Madan and Schoutens (2018), Munnix et al. (2014), Sidi (2020), Yan et al. (2019), Yaros and Imielinski (2015), and Yeh (2020), Paparrizos et al. (2020), Zhu et al. (2020), Suarez et al. (2020), Mishra et al. (2020), Li et al. (2020c), Zhang et al. (2020c), Ontanon (2020), Costa et al. (2020), Edelmann et al. (2019)
  - combine observations with expert views: Czasonis et al. (2020), Davis and Lleo (2016), Meucci (2009), Yu et al. (2020), Rosen and Saunders (2016), Rosen (2015b), Yu et al. (2014), Kolm and Ritter (2020a)
  - agent based modeling: Raman et al. (2019), Haldane and Turrell (2018), Beikirch et al. (2018), Bouchaud (2018), Chakrabarti et al. (2019),Mizuta (2019), Lamperti et al. (2018), Lussange et al. (2020), Cramer and Trimborn (2019),Angione et al. (2020), Raman and Leidner (2019), Byrd (2019), Dahlke et al. (2020), Zhang et al. (2020b), Tran et al. (2020), Shiono (2020), Fievet and Sornette (2018)
  - scenario generation (relevant references on next slide)
  - construction of stress scenarios (relevant references on next slide)

Methods to generate synthetic data (cont.)

- Methods to deliver consistent and realistic synthetic data also include
  - scenario generation (including factor-based and reduced form modeling): Golub et al. (2018), Dunkler et al. (2014), Engle and Roussellet (2016), Buehler et al. (2020b), Buehler et al. (2020a), Czasonis et al. (2020), Iwana and Uchida (2020), Kondratyev and Schwarz (2020), Yoon (2020), Kaut (2020), Fox and Jha (2019), Hazarika et al. (2020), Waller (2020), Lam and Li (2020), Seib et al. (2020), Kang et al. (2020), Oh et al. (2020), Kondratyev et al. (2020), Ni et al. (2020b), Koochali et al. (2020), Pesenti et al. (2020), Ponomareva et al. (2015), Rosen (2015a), Sironi (2015), Tang et al. (2018), van Beek (2020), Facchinato and Pola (2014), Gosling (2010), Kaut and Lium (2015), Kegel et al. (2017), Liu (2015)
  - construction of stress scenarios: Glasserman et al. (2015), Kopeliovich et al. (2015), Grundke (2012), Rosen and Saunders (2016), Skoglund (2019), Rebonato (2018), Bilgili et al. (2017), Denev and Mutnikas (2016b), Ardia and Meucci (2015), Flood and Korenko (2015), Meucci (2012), Overbeck (2012). Siddique et al. (2019), Traccucci et al. (2019), Ruenzi et al. (2020),Baes and Schaanning (2020), Engle (2020), Cohort et al. (2020), Breuer and Summer (2020), Albanese et al. (2020), Aste (2020), Pesenti et al. (2020), Pesenti et al. (2019)

Challenge: incorporate causality in $\mathbb{ML}$

- When humans rationalize the world, we often think in terms of cause and effect. Causal inference is a statistical tool for $\mathbb{ML}$ that enables analysis of $\mathbb{ML}$ along the same lines.
- Knight (2019): interview with Yoshua Bengio: *"It's a big thing to integrate [causality] into AI. Current approaches to ML assume that the trained AI system will be applied on the same kind of data as the training data. In real life it is often not the case"*.
- Jacob et al. (2020a), Cui et al. (2020), Guo et al. (2020a): surveys of both traditional and frontier methods in learning causality and connections between causality and $\mathbb{ML}$
- Yao et al. (2020b), Moraffah et al. (2020): reviews of causal inference methods under potential outcome framework, and of causal interpretable models (aiming to answer causal questions)
- Mastakouri et al. (2020): identify direct and indirect causes on time series and provide necessary and sufficient conditions in presence of latent variables

Challenge: incorporate causality in $\mathbb{ML}$ (cont.)

- Gnecco et al. (2020), Pearl and Mackenzie (2018), Pearl (2019): tools of causal inference
- Baldi and Shahbaba (2020): causality within Bayesian statistical framework
- Wang et al. (2020b), Huang et al. (2020a), Wong (2020), Chen et al. (2020b): methods and implementations for causal $\mathbb{ML}$
- Huang et al. (2020b), da Costa and Dasgupta (2020), Saggioro et al. (2020): detect causal relationships identified from time series
- Little and Badawy (2020): causal bootstrapping to resample observational data such that, if an interventional relationship is identified from that data, new data representing that relationship can be simulated
- Based on such resampled data, $\mathbb{ML}$-based forecasting is both statistically powerful and causally robust
- Lowe et al. (2020): Amortized Causal Discovery leverages shared dynamics to learn to infer causal relations from time-series data.
- Parafita and Vitria (2020): causal inference using Deep Causal Graphs

Challenge: interpretability and explainability

- Question: *If a machine learning model performs well, why not just trust the model and ignore why it made a certain decision?*
- Kilburn (2020a): based on survey of practitioners, this question becomes especially relevant when using $\mathbb{ML}$ in $\mathbb{QWIM}$
- Stakeholders and regulators require significant levels of interpretability and/or explainability for $\mathbb{ML}$ in $\mathbb{QWIM}$
- Explainability and interpretability are often used interchangeably, although they are not similar
- Zhou (2019): *interpretability* (describing HOW $\mathbb{ML}$ arrived at its conclusions) versus *explainability* (answering WHY)
- Molnar (2020a), Choi (2020), Samek et al. (2020), Stewart (2020), Kim (2019), Biecek and Burzykowski (2021), Doshi-Velez and Kim (2017): surveys and methods for $\mathbb{ML}$ interpretability
- Molnar et al. (2020): pitfalls to avoid when interpreting $\mathbb{ML}$

Interpretability

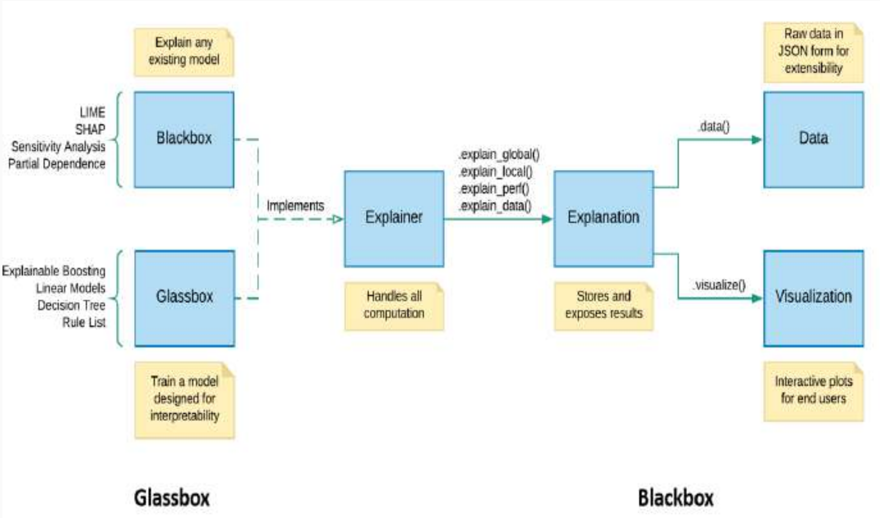
- Molnar (2020b): Methods to quantify interpretability:
  - Model-agnostic methods
    - Global and Local surrogate models
    - Feature importance and interaction
    - Individual Coditional Expectations
    - Shapley Additive exPlanations SHAP
    - Scoped rules (anchors)
    - Partial Dependence Plot PDP
    - Accumulated Local Effects Plot ALEP
    - Merging Path Plot MPP
    - Break Down Plot BDP
    - Permutational Variable Importance Plot PVIP
    - Cateris Paribus Plot CPP
  - Example-based explanations
    - Counterfactual explanations
    - Adversarial examples
    - Prototypes and criticisms
    - Influential instances
- Hall (2020): comprehensive list of resources on interpretability

Interpretability (cont.)

- Ribeiro et al. (2016a), Ribeiro et al. (2016b): LIME ("Local Interpretable Model-Agnostic Explanations") used to fit local surrogate models:
  - Choose instance of interest for which to have an explanation
  - Perturb dataset and get black box predictions for these new points.
  - Weight new samples by their proximity to instance of interest.
  - Fit weighted, interpretable model on dataset with variations.
  - Explain prediction by interpreting the local model.
- Ribeiro et al. (2018): "Anchors" to address LIME weaknesses
  - Inability to know how widely one can apply a "local" explanation
  - Dependence on assumptions of linearity
  - Dependence on rather vague concepts of closeness/distance
- Ribeiro et al. (2018): further enhancements "bLIMEy" and "LIMEtree"
- Linden et al. (2019): argue that global importance introduced by LIME does not reliably represent the model's global behavior, and introduce Global Aggregations of Local Explanations (GALE)

Software packages for interpretability

- While there are many such packages, we mention two software suites which have proven quite powerful: InterpretML and DALEX
- Nori et al. (2019): InterpretML uses glassbox and blackbox

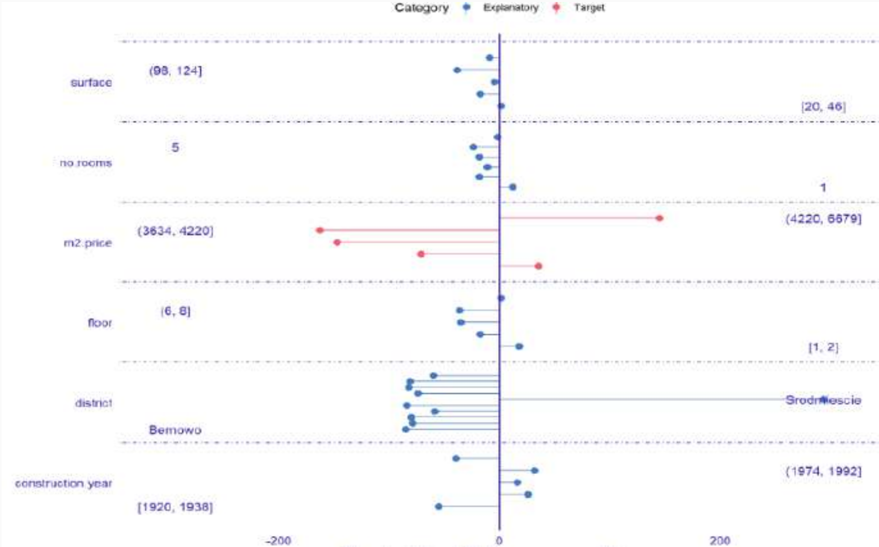


Source: Nori et al. (2019)

Software packages for interpretability (cont.)

- Biecek and Burzykowski (2021), Biecek (2018): Descriptive mAchine Learning EXplanations DALEX suite

Figure 2: Champion (left) vs. Challenger (right)



Source: Biecek and Maksymiuk (2019)

Explainability

- According to DARPA, eXplainable artificial intelligence $\mathbb{XAI}$ aims to "*produce more explainable models, while maintaining a high level of learning performance (prediction accuracy); and enable human users to understand, appropriately, trust, and effectively manage the emerging generation of artificially intelligent partners*"
- Biecek and Burzykowski (2021), Choi (2020), Caruana et al. (2020), Ignatiev (2020), Adadi and Berrada (2018), Danilevsky et al. (2020), Camburu (2020), Dosilovic et al. (2018), Samek et al. (2019), Arya et al. (2020), Heuillet et al. (2020), Xie et al. (2020), Puiutta and Veith (2020): surveys on current state-of-the-art, developments, and challenges for $\mathbb{XAI}$, including $\mathbb{XDL}$ (explainable deep learning) and $\mathbb{XRL}$ (explainable reinforcement learning)
- Arrieta et al. (2019): concept of Responsible Artificial Intelligence for large-scale implementation of $\mathbb{AI}$ organizations with fairness, model explainability and accountability at its core
- Schneider and Handali (2019): concept of personalized explanation in $\mathbb{XAI}$ based on three key explanation properties: complexity, decision information and presentations

Explainability (cont.)

- Spinner et al. (2020), Choo and Liu (2018), Samek et al. (2017), Zhang and Zhu (2018): visual analytics and information visualization within context of $\mathbb{XAI}$
- Horel and Giesecke (2020), Horel and Giesecke (2018), Horel and Giesecke (2019): computationally efficient significance tests for $\mathbb{ML}$ features and feature interactions of any order in a hierarchical manner; generates a model-free notion of feature importance
- Horel et al. (2019): $\mathbb{XML}$ applied to wealth management compliance
- Roscher et al. (2020) : survey on $\mathbb{XML}$ combined with domain knowledge
- Chen et al. (2020c): significance, relevance and explainability for $\mathbb{ML}$, from econometrics and financial data science perspectives
- Chen et al. (2020c) and Hoepner et al. (2021): practitioner framework for Adaptive Explainable Neural Networks within context of quantitative finance

Explainability (cont.)

- Virgolin et al. (2019), Udrescu et al. (2020), Udrescu and Tegmark (2020): symbolic regression can improve $\mathbb{ML}$ explainability
- Marcinkevics and Vogt (2020): emphasise divide between interpretability and explainability and illustrate these two different research directions with concrete examples of state-of-the-art.
- Spinner et al. (2020): framework for interactive $\mathbb{XAI}$ to allow users to
  - understand $\mathbb{ML}$ models
  - diagnose model limitations using different $\mathbb{XAI}$ methods
  - refine and optimize the models
- Framework combines iterative $\mathbb{XAI}$ pipeline with eight global monitoring and steering mechanisms, including quality monitoring, provenance tracking, model comparison, and trust building

Challenge: fairness, bias, and data privacy

- Mireshghallah et al. (2020), De Cristofaro (2020): review of data privacy challenges and possible countermeasures
- Rodriguez-Barroso et al. (2020), Triastcyn and Faltings (2020), Shen et al. (2021) Ntoutsi et al. (2020): methodologies and software toolsfor preserving data privacy
- Term “bias” is used in conjunction with $\mathbb{ML}$ in many different contexts, and with many different meanings
- Hellstrom et al. (2020), Ntoutsi et al. (2020): taxonomy and analysis of how different types of biases are connected and depend on each other
- Saleiro et al. (2020): mitigation frameworks for bias and fairness when building and deploying data science systems
- Sharma et al. (2019): CERTIFAI - Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of $\mathbb{AI}$ models
- Segal et al. (2020), Dwork et al. (2020), Makhlouf et al. (2020), Oneto and Chiappa (2020), Ruf et al. (2020), del Barrio et al. (2020): concepts, metrics and procedures to assess and ensure fairness in $\mathbb{ML}$

Challenge: tuning hyperparameters

- $\mathbb{ML}$ hyperparameters versus parameters
  - $\mathbb{ML}$ hyperparameters define model configuration (activation functions, optimization methods, training/testing sets)
  - $\mathbb{ML}$ model parameters: learned during model training
- Bakhteev and Strijov (2020), Yu and Zhu (2020), Yang and Shami (2020), Feurer and Hutter (2019): Hyperparameter search can be performed on a predefined grid, via rules-of-thumb, random search, systematic (such as automated $\mathbb{ML}$ and multiobjective optimization)
- van Rijn and Hutter (2018): *"Given an algorithm, what are its most important hyperparameters, and what are typically good values for these?"*
- Probst et al. (2019): measuring tunability of hyperparameters
- Nystrup et al. (2020), Feurer and Hutter (2019): systematic hyperparameter tuning delivers better investment portfolios
- Parameters may also be specialized to tasks, such as finding optimal number of clusters: Kawamoto and Kabashima (2017b), Kawamoto and Kabashima (2017a), Riolo et al. (2017)

Challenge: tuning hyperparameters (cont.)

- Bezrukavnikov and Linder (2021), Halvari et al. (2020), Milutinovic et al. (2020), Tuggener et al. (2019), Hanussek et al. (2020), Zoller and Huber (2020): recent advances in automated hyperparameter search (automated $\mathbb{ML}$)
- Hyperparamer optimization HPO frameworks: HyperOpt, Spearmint, SMAC, Autotune, Vizier, Katib, Tune, Hyperband, GPyOpt
- Various choice of parameter-sampling algorithms: Gaussian processes in Spearmint and GPyOpt, tree-structured Parzen estimator in HyperOpt, random forests in SMAC
- Recent frameworks (Vizier, Katib, Tune) support pruning algorithms, which stop unpromising trials prematurely to speed up the exploration
- Klein et al. (2019): meta-surrogate benchmarking for HPO, to generate inexpensive and realistic scenarios
- Jomaa et al. (2019): HPO as a sequential decision problem (which hyperparameter to test next) solved with reinforcement learning
- Akiba et al. (2019): Optuna: A Next-generation HPO Framework


100
90
80
70
60
50
40
30
20
10
0
100
150
200
250
300
350
400

References

Abolghasemi, M., Hyndman, R. J., Tarr, G., and Bergmeir, C. (2019). "Machine learning applications in time series hierarchical forecasting." In: arXiv e-Print.
Abouseir, A., Le Manach, A., El Mennaoui, M., and Zheng, B. (2020). "Integration of Macroeconomic Data into Multi-Asset Allocation with Machine Learning Techniques." In: SSRN e-Print.
Ada, S. E., Ugur, E., and Akin, H. L. (2019). "Generalization in Transfer Learning." In: arXiv e-Print.
Adadi, A. and Berrada, M. (2018). "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)." In: IEEE Access: Practical Iinnovations, Open Solutions 6, pp. 52138–52160.
Adammer, P. and Schussler, R. A. (2020). "Forecasting the Equity Premium: Mind the News!" In: Review of Finance, Early View.
Adolfsson, A., Ackerman, M., and Brownstein, N. C. (2019). "To cluster, or not to cluster: An analysis of clusterability methods." In: Pattern Recognition 88, pp. 13–26.
Agarwal, V. (2020). "Unearthing Alpha in Financial Annual Reports." In: SSRN e-Print.
Aggarwal, C. C. (2018). Machine Learning for Text. Springer International Publishing. 493 pp.
Aguilera, A. C., Artes-Rodriguez, A., Perez-Cruz, F., and Olmos, P. M. (2020). "Robust Sampling in Deep Learning." In: arXiv e-Print.
Ahern, I., Noack, A., Guzman-Nateras, L., Dou, D., Li, B., and Huan, J. (2019). "NormLime: A New Feature Importance Metric for Explaining Deep Neural Networks." In: arXiv e-Print.
Ahuja, K., Zame, W. R., and Schaar, M. van der (2018). "Piecewise Approximations of Black Box Models for Model Interpretation." In: arXiv e-Print.
Aich, S., Al-Absi, A. A., Hui, K. L., and Sain, M. (2019). "Prediction of Quality for Different Type of Wine based on Different Feature Sets Using Supervised Machine Learning Techniques." In: 21st International Conference on Advanced Communication Technology (ICACT).
Akiba, T., Sano, S., Yanase, T., Ohta, T., and Koyama, M. (2019). "Optuna: A Next-generation Hyperparameter Optimization Framework." In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. KDD '19. Anchorage, AK, USA: ACM, pp. 2623–2631.

Albanese, C., Crepey, S., and Iabichino, S. (2020). "Reverse Stress Testing." In: SSRN e-Print.

Alberico, S., Coche, J., Sahakyan, V., and Zulaica, O. (2018). "Regime identification for sovereign bond portfolio construction." In: Advances in the practice of public investment management: portfolio modelling, performance attribution and governance. Ed. by N. Bulusu, J. Coche, A. Reveiz, F. Rivadeneyra, V. Sahakyan, and G. Yanou. Springer International Publishing, pp. 247–274.

Allen, D. and Satchell, S. (2014). The Four Horsemen: Heavy-tails, Negative Skew, Volatility Clustering, Asymmetric Dependence. Tech. rep. DP-004. University of Sydney.

Allende, H. and Valle, C. (2017). "Ensemble methods for time series forecasting." In: Claudio Moraga: A Passion for Multi-Valued Logic and Soft Computing. Ed. by R. Seising and H. Allende-Cid. Vol. 349. Springer International Publishing, pp. 217–232.

Alonso, M. N., Batres-Estrada, G., and Moulin, A. (2019). "Deep Learning in Finance: Prediction of Stock Returns with Long Short-Term Memory Networks." In: Big data and machine learning in quantitative investment. Wiley, pp. 251–277.

Alonso, M. N. i, Batres-Estrada, G., and Moulin, A. (2020). "Deep Learning for Equity Time Series Prediction." In: SSRN e-Print.

Alsabah, H., Capponi, A., Ruiz Lacedelli, O., and Stern, M. (2021). "Robo-Advising: Learning Investors Risk Preferences via Portfolio Choices." In: Journal of Financial Econometrics, Early View.

Alves, P., Duarte, L., and Mota, L. da (2018). "Detecting chaos and predicting in Dow Jones Index." In: Chaos, Solitons and Fractals 110, pp. 232–238.

Amarasinghe, K., Kenney, K., and Manic, M. (2018). "Toward explainable deep neural network based anomaly detection." In: 11th International Conference on Human System Interaction (HSI). IEEE, pp. 311–317.

Amel-Zadeh, A., Calliess, J.-P., Kaiser, D., and Roberts, S. (2020). "Machine Learning-Based Financial Statement Analysis." In: SSRN e-Print.

Aminikhanghahi, S. and Cook, D. (2017). "A survey of methods for time series change point detection." In: Knowledge and Information Systems 51(2), pp. 339–367.

Ammann, M. and Verhofen, M. (2006). "The effect of market regimes on style allocation." In: Financial Markets and Portfolio Management 20(3), pp. 309–337.

Andreini, P., Izzo, C., and Ricco, G. (2020). "Deep Dynamic Factor Models." In: arXiv e-Print.

Ang, A. and Timmermann, A. (2012). "Regime Changes and Financial Markets." In: Annual Review of Financial Economics 4(1), pp. 313–337.

Angione, C., Silverman, E., and Yaneske, E. (2020). "Using Machine Learning to Emulate Agent-Based Simulations." In: arXiv e-Print.

Antoncic, M., Bekaert, G., Rothenberg, R. V., and Noguer, M. (2020). "Sustainable Investment - Exploring the Linkage between Alpha, ESG, and SDG's." In: SSRN e-Print.

Aparicio, D. and Lopez de Prado, M. (2018). "How Hard Is It to Pick the Right Model? MCS and backtest overfitting." In: Algorithmic Finance 7, pp. 53–61.

Ardia, D. and Meucci, A. (2015). "Stress testing in non-normal markets via entropy pooling." In: Risk.

Arik, S. O. and Pfister, T. (2019). "Attention-Based Prototypical Learning Towards Interpretable, Confident and Robust Deep Neural Networks." In: arXiv e-Print.

Arnott, R. D., Harvey, C. R., and Markowitz, H. (2019). "A backtesting protocol in the era of machine learning." In: The Journal of Financial Data Science 1(1), pp. 64–74.

Arras, L., Arjona-Medina, J., Widrich, M., Montavon, G., Gillhofer, M., Muller, K.-R., Hochreiter, S., and Samek, W. (2019). "Explaining and interpreting LSTMs." In: Explainable AI: interpreting, explaining and visualizing deep learning. Ed. by W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Muller. Vol. 11700. Springer International Publishing, pp. 211–238.

Arrieta, A. B., Diaz-Rodriguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., and Herrera, F. (2019). "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI." In: arXiv e-Print.

Arya, V., Bellamy, R. K. E., Chen, P.-Y., Dhurandhar, A., Hind, M., Hoffman, S. C., Houde, S., Liao, Q. V., Luss, R., Mojsilovic, A., Mourad, S., Pedemonte, P., Raghavendra, R., Richards, J. T., Sattigeri, P., Shanmugam, K., Singh, M., Varshney, K. R., Wei, D., and Zhang, Y. (2020). "AI Explainability 360: An Extensible Toolkit for Understanding Data and Machine Learning Models." In: Journal of Machine Learning Research 21(130), pp. 1–6.

Assefa, S., Dervovic, D., Mahfouz, M., Balch, T., Reddy, P., and Veloso, M. (2019). "Generating Synthetic Data in Finance: Opportunities, Challenges and Pitfalls." In: NeurIPS'19 Workshop on Robust AI in Financial Services.

Aste, T. (2020). "Stress testing and systemic risk measures using multivariate conditional probability." In: arXiv e-Print.

Athanasopoulos, G., Gamakumara, P., Panagiotelis, A., Hyndman, R. J., and Affan, M. (2019). "Hierarchical Forecasting." In: Macroeconomic Forecasting in the Era of Big Data. Springer International Publishing, pp. 689–719.

Athanasopoulos, G. and Kourentzes, N. (2020). On the evaluation of hierarchical forecasts. Tech. rep. Monash University.

Atiya, A. F. (2020). "Why does forecast combination work so well?" In: International Journal of Forecasting 36(1), pp. 197–200.

Audrino, F., Sigrist, F., and Ballinari, D. (2020). "The impact of sentiment and attention measures on stock market volatility." In: International Journal of Forecasting 36(2), pp. 334–357.

Avigdor-Elgrabli, N., Libov, A., Viderman, M., and Wolff, R. (2018). "Interpreting Complex Regression Models." In: arXiv e-Print.

Babii, A., Ghysels, E., and Striaukas, J. (2020). "Machine learning time series regressions with an application to nowcasting." In: arXiv e-Print.

Bacciu, D., Errica, F., Micheli, A., and Podda, M. (2020). "A gentle introduction to deep learning for graphs." In: Neural Networks 129, pp. 203–221.

Bachoc, F., Gamboa, F., Loubes, J.-M., and Risser, L. (2018). "Entropic Variable Boosting for Explainability and Interpretability in Machine Learning." In: arXiv e-Print.

Baes, M. and Schaanning, E. (2020). "Reverse Stress Testing: Scenario Design for Macroprudential Stress Tests." In: SSRN e-Print.

Bahra, B. and Thukral, L. (2020). "ESG in Global Corporate Bonds: The Analysis Behind the Hype." In: The Journal of Portfolio Management, Early View.

Bai, L., Cui, L., Xu, L., Wang, Y., Zhang, Z., and Hancock, E. R. (2019). "Entropic Dynamic Time Warping Kernels for Co-evolving Financial Time Series Analysis." In: arXiv e-Print.

Bailey, D. H., Borwein, J. M., Lopez de Prado, M., and Zhu, Q. J. (2014). "Pseudo-Mathematics and Financial Charlatanism: The Effects of Backtest Overfitting on Out-of-Sample Performance." In: Notices of the American Mathematical Society 61(5), pp. 458–471.

Bailey, D. H., Borwein, J. M., and de Prado, M. L. (2017). "Stock Portfolio Design and Backtest Overfitting." In: Journal of Investment Management 15(1).

Baitinger, E. and Papenbrock, J. (2017a). "Interconnectedness Risk and Active Portfolio Management." In: Journal of Investment Strategies 6(2), pp. 63–90.

Baitinger, E. and Papenbrock, J. (2017b). "Interconnectedness risk and active portfolio management: the information-theoretic perspective." In: The Journal of Network Theory in Finance 3 (4), pp. 25–47.

Bakhteev, O. Y. and Strijov, V. V. (2020). "Comprehensive analysis of gradient-based hyperparameter optimization algorithms." In: Annals of Operations Research 289(1), pp. 51–65.

Balasubramanian, K., Sundaresh, H., Wu, D., and Kearns, K. (2019). "Do stocks stalk other stocks in their complex network? A complex networks approach to stock market dynamics." In: The Journal of Financial Data Science 1 (2), pp. 35–54.

Baldi, P. and Shahbaba, B. (2020). "Bayesian Causality." In: The American Statistician 74(3), pp. 249–257.

Baltas, N. and Karyampas, D. (2018). "Forecasting the equity risk premium: The importance of regime-dependent evaluation." In: Journal of Financial Markets 38(March), pp. 83–102.

Bandara, K., Shi, P., Bergmeir, C., Hewamalage, H., Tran, Q., and Seaman, B. (2019). "Sales Demand Forecast in E-commerce using a Long Short-Term Memory Neural Network Methodology." In: Proceedings of International Conference on Neural Information Processing ICONIP.

Bao, W., Yue, J., and Rao, Y. (2017). "A deep learning framework for financial time series using stacked autoencoders and long-short term memory." In: PLOS ONE 12(7), e0180944+.

Barandas, M., Folgado, D., Fernandes, L., Santos, S., Abreu, M., Bota, P., Liu, H., Schultz, T., and Gamboa, H. (2020). "TSFEL: time series feature extraction library." In: SoftwareX 11, p. 100456.

Barras, L. (2019). "A large-scale approach for evaluating asset pricing models." In: Journal of Financial Economics 134(3), pp. 549–569.

Barras, L., Scaillet, O., and Wermers, R. (2019). "Reassessing False Discoveries in Mutual Fund Performance: Skill, Luck, or Lack of Power? A Reply." In: SSRN e-Print.

Barras, L., Scaillet, O., and Wermers, R. (2010). "False discoveries in mutual fund performance: Measuring luck in estimated alphas." In: Journal of Finance 65, pp. 179–216.

Bartram, S. M., Branke, J., and Motahari, M. (2020). "Artificial Intelligence in Asset Management." In: CFA Research Foundation Literature Reviews.

Bashar, M. A. and Nayak, R. (2020). "TAnoGAN: Time Series Anomaly Detection with Generative Adversarial Networks." In: arXiv e-Print.

Bashchenko, O. and Marchal, A. (2020). "Deep learning for asset bubbles detection." In: SSRN e-Print.

Bauer, A., Zufle, M., Grohmann, J., Schmitt, N., Herbst, N., and Kounev, S. (2020a). "An Automated Forecasting Framework based on Method Recommendation for Seasonal Time Series." In: Proceedings of the ACM/SPEC International Conference on Performance Engineering.

Bauer, A., Zufle, M., Herbst, N., Kounev, S., and Curtef, V. (2020b). "Telescope: An Automatic Feature Extraction and Transformation Approach for Time Series Forecasting on a Level-Playing

Field." In: Proceedings of the 36th International Conference on Data Engineering (ICDE), pp. 1902–1905.

Begusic, S. and Kostanjcar, Z. (2019). "Cluster-Based Shrinkage of Correlation Matrices for Portfolio Optimization." In: 11th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE, pp. 301–305.

Beikirch, M., Cramer, S., Frank, M., Otte, P., Pabich, E., and Trimborn, T. (2018). "Simulation of Stylized Facts in Agent-Based Computational Economic Market Models." In: arXiv e-Print.

Bektic, D., Hachenberg, B., and Schiereck, D. (2020). "Factor-based investing in government bond markets: a survey of the current state of research." In: Journal of Asset Management 21, pp. 94–105.

Benhamou, E. (2018). "Kalman filter demystified: from intuition to probabilistic graphical model to real case in financial markets." In: arXiv e-Print.

Benhamou, E., Saltiel, D., Ungari, S., and Mukhopadhyay, A. (2020). "Time your hedge with Deep Reinforcement Learning." In: arXiv e-Print.

Benhamou, E., Saltiel, D., Ungari, S., and Mukhopadhyay, A. (2021). "Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning." In: SSRN e-Print.

Benidis, K., Rangapuram, S. S., Flunkert, V., Wang, B., Maddix, D., Turkmen, C., Gasthaus, J., Bohlke-Schneider, M., Salinas, D., Stella, L., Callot, L., and Januschowski, T. (2020). "Neural forecasting: Introduction and literature overview." In: arXiv e-Print.

Bernhart, G., Hocht, S., Neugebauer, M., Neumann, M., and Zagst., R. (2011). "Asset Correlations in Turbulent Markets and the Impact of Different Regimes on Asset Management." In: Asia-Pacific Journal of Operational Research 28(01), pp. 1–23.

Bezrukavnikov, O. and Linder, R. (2021). "A Neophyte With AutoML: Evaluating the Promises of Automatic Machine Learning Tools." In: arXiv e-Print.

Bianchi, D., Buchner, M., and Tamoni, A. (2021). "Bond Risk Premiums with Machine Learning." In: The Review of Financial Studies, Early View.

Biecek, P. (2018). "DALEX: explainers for complex predictive models." In: Journal of Machine Learning Research 19, pp. 1–5.
Biecek, P. and Burzykowski, T. (2021). Explanatory Model Analysis: Explore, Explain, and Examine Predictive Model. Chapman and Hall/CRC. 310 pp.
Biecek, P. and Maksymiuk, S. (2019). "DALEXtra." In: Journal of Machine Learning Research.
Bilgili, M. S., Ferconi, M., and Ulitsky, A. (2017). "Stress Hedging in Portfolio Construction." In: Risk.
Billio, M., Costola, M., Hristova, I., Latino, C., and Pelizzon, L. (2020). Inside the ESG ratings: (Dis)agreement and performance. Tech. rep. SAFE.
Billio, M. and Varotto, S. (2020). "A New World Post COVID-19: Lessons for Business, the Finance Industry and Policy Makers." In: SSRN e-Print.
Blazquez-Garcia, A., Conde, A., Mori, U., and Lozano, J. A. (2020). "A review on outlier/anomaly detection in time series data." In: arXiv e-Print.
Blitz, D. and Van Vliet, P. (2009). "Dynamic Strategic Asset Allocation: Risk and Return Across Economic Regimes." In: SSRN e-Print.
Bloch, D. A. (2020a). "Machine Learning: Models And Algorithms." In: SSRN e-Print.
Bloch, D. A. (2020b). "Recipe for Quantitative Trading with Machine Learning." In: SSRN e-Print.
BNP Paribas (2019). The ESG Global Survey 2019. Tech. rep. BNP Paribas.
Boehmke, B. and Greenwell, B. (2019). Hands-On Machine Learning with R. Taylor & Francis. 488 pp.
Bolhuis, M. and Rayner, B. (2020). "Deus Ex Machina? A Framework for Macro Forecasting with Machine Learning." In: IMF Working Papers (20).
Bolon-Canedo, V. and Alonso-Betanzos, A. (2019). "Ensembles for feature selection: A review and future trends." In: Information Fusion 52, pp. 1–12.
Borghi, R. and Rossi, G. D. (2020). "The Artificial Intelligence Approach to Picking Stocks." In: Machine Learning for Asset Management: New Developments and Financial Applications. Wiley, pp. 115–166.

Botchkarev, A. (2019). "A new typology design of performance metrics to measure errors in machine learning regression algorithms." In: Interdisciplinary Journal of Information, Knowledge, and Management 14, pp. 045–076.

Bouchaud, J.-P. (2018). "Agent-based models for market impact and volatility." In: Handbook of Computational Economics. Vol. 4. Elsevier. Chap. 7, pp. 393–416.

Boudt, K., Cela, M., and Simaan, M. (2020). "In Search of Return Predictability: Application of Machine Learning Algorithms in Tactical Allocation." In: Machine Learning for Asset Management: New Developments and Financial Applications. Ed. by E. Jurczenko. Wiley, pp. 35–73.

Bouri, E., Gupta, R., Wong, W.-K., and Zhu, Z. (2018). "Is wine a good choice for investment?" In: Pacific-Basin Finance Journal 51, pp. 171–183.

Braei, M. and Wagner, S. (2020). "Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art." In: arXiv e-Print.

Bramson, A. and Vandermarliere, B. (2016). "Benchmarking Measures of Network Influence." In: Scientific Reports 6(1).

Breuer, T. and Summer, M. (2020). "Systematic stress tests on public data." In: Journal of Banking & Finance 118, pp. 105886+.

Briere, M. and Signori, O. (2012). "Inflation-Hedging Portfolios: Economic Regimes Matter." In: The Journal of Portfolio Management 38(4), pp. 43–58.

Brigato, L. and Iocchi, L. (2020). "A Close Look at Deep Learning with Small Data." In: arXiv e-Print.

Brown, A. and Petrik, M. (2018). "Interpretable Reinforcement Learning with Ensemble Methods." In: arXiv e-Print.

Brunel, J. L. P. (2011). "Goal-Based Wealth Management in Practice." In: The Journal of Wealth Management 14(3), pp. 17–26.

Bryzgalova, S., Huang, J., and Julliard, C. (2020). "Bayesian solutions for the factor zoo: we just ran two quadrillion models." In: SSRN e-Print.

Buehler, H., Gonon, L., Teichmann, J., and Wood, B. (2019). "Deep hedging." In: Quantitative Finance 19 (8), pp. 1271–1291.

Buehler, H., Horvath, B., Lyons, T., Arribas, I. P., and Wood, B. (2020a). "A Data-driven Market Simulator for Small Data Environments." In: arXiv e-Print.

Buehler, H., Horvath, B., Lyons, T., Arribas, I. P., and Wood, B. (2020b). "Generating financial markets with signatures." In: SSRN e-Print.

Bugata, P. and Drotar, P. (2019). "Weighted nearest neighbors feature selection." In: Knowledge-Based Systems 163, pp. 749–761.

Bulusu, S., Kailkhura, B., Li, B., Varshney, P. K., and Song, D. (2020). "Anomalous Instance Detection in Deep Learning: A Survey." In: arXiv e-Print.

Burkov, A. (2019). The Hundred-Page Machine Learning Book. True Positive Inc. 160 pp.

Byrd, D. (2019). "Explaining Agent-Based Financial Market Simulation." In: arXiv e-Print.

Byrd, D., Bajaj, S., and Balch, T. (2019). "Fund Asset Inference Using Machine Learning Methods: What's in That Portfolio?" In: The Journal of Financial Data Science 1 (3), pp. 98–107.

Cai, F., Le-Khac, N.-A., and Kechadi, T. (2016). "Clustering Approaches for Financial Data Analysis: a Survey." In: arXiv e-Print.

Cai, J., Luo, J., Wang, S., and Yang, S. (2018). "Feature selection in machine learning: A new perspective." In: Neurocomputing 300, pp. 70–79.

Camburu, O.-M. (2020). "Explaining Deep Neural Networks." In: arXiv e-Print.

Campos, L. (2018). "Deep Reinforcement Trading." In: QuantDare.

Cao, J., Chen, J., Hull, J., and Poulos, Z. (2021). "Deep Hedging of Derivatives Using Reinforcement Learning." In: The Journal of Financial Data Science, Early View.

Cao, J., Chen, J., Hull, J. C., and Poulos, Z. (2020). "Deep Hedging of Derivatives Using Reinforcement Learning." In: SSRN e-Print.

Carbonneau, A. (2020). "Deep Hedging of Long-Term Financial Derivatives." In: arXiv e-Print.

Carbonneau, A. and Godin, F. (2020). "Equal risk pricing of derivatives with deep hedging." In: Quantitative Finance, Early View.

Caruana, R., Lundberg, S., Ribeiro, M. T., Nori, H., and Jenkins, S. (2020). "Intelligible and Explainable Machine Learning." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Cavallo, M. and Demiralp, C. (2018). "Clustrophile 2: Guided Visual Clustering Analysis." In: arXiv e-Print.

Cepoi, C.-O. (2020). "Asymmetric dependence between stock market returns and news during COVID-19 financial turmoil." In: Finance Research Letters 36, p. 101658.

Cerqueira, V., Torgo, L., and Soares, C. (2019). "Machine Learning vs Statistical Methods for Time Series Forecasting: Size Matters." In: arXiv e-Print.

Chakrabarti, A. S., Pichl, L., and Kaizoji, T., eds. (2019). Network Theory and Agent-Based Modeling in Economics and Finance. Springer Singapore. 458 pp.

Chalapathy, R. and Chawla, S. (2019). "Deep Learning for Anomaly Detection: A Survey." In: arXiv e-Print.

Chalapathy, R., Khoa, N. L. D., and Chawla, S. (2020). "Robust Deep Learning Methods for Anomaly Detection." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Chang, C.-C., Lin, Z.-T., Koc, W.-W., Chou, C., and Huang, S.-H. (2016). "Affinity propagation clustering for intelligent portfolio diversification and investment risk reduction." In: 7th International Conference on Cloud Computing and Big Data (CCBD). IEEE, pp. 145–150.

Charte, D., Charte, F., Garcia, S., Jesus, M. J. del, and Herrera, F. (2018). "A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines." In: Information Fusion 44, pp. 78–96.

Chazal, F. and Michel, B. (2017). "An introduction to Topological Data Analysis: fundamental and practical aspects for data scientists." In: arXiv e-Print.

Chen, A. Y. and Zimmermann, T. (2020). "Open Source Cross-Sectional Asset Pricing." In: SSRN e-Print.

Chen, C.-C., Huang, H.-H., and Chen, H.-H. (2020a). "NLP in FinTech Applications: Past, Present and Future." In: arXiv e-Print.

Chen, H., Harinen, T., Lee, J.-Y., Yung, M., and Zhao, Z. (2020b). "CausalML: Python Package for Causal Machine Learning." In: arXiv e-Print.

Chen, J., Vaughan, J., Nair, V., and Sudjianto, A. (2020c). "Adaptive explainable neural networks AxNNs." In: SSRN e-Print.

Chen, M. and Mussalli, G. (2020). "An integrated approach to quantitative ESG investing." In: The Journal of Portfolio Management 46(3), pp. 65–74.

Cheng, I.-H. (2020). "Volatility Markets Underreacted to the Early Stages of the COVID-19 Pandemic." In: Review of Asset Pricing Studies, Early View.

Chernozhukov, V., Hansen, C., and Spindler, M. (2015). "Valid Post-Selection and Post-Regularization Inference: An Elementary, General Approach." In: Annual Review of Economics 7(1), pp. 649–688.

Chhabra, A. B. (2015). The Aspirational Investor: Taming the Markets to Achieve Your Life's Goals. HarperBusiness. 245 pp.

Chib, S. and Zeng, X. (2020). "Which factors are risk factors in asset pricing? A model scan framework." In: Journal of Business & Economic Statistics 38(4), pp. 771–783.

Chib, S., Zeng, X., and Zhao, L. (2020). "On Comparing Asset Pricing Models." In: Journal of Finance 75(11), pp. 551–577.

Chincoli, F. and Guidolin, M. (2017). "Linear and nonlinear predictability in investment style factors: multivariate evidence." In: Journal of Asset Management 18, pp. 476–509.

Cho, M. and Hegde, C. (2019). "Reducing the search space for hyperparameter optimization using group sparsity." In: ICA 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICA). IEEE, pp. 3627–3631.

Choi, J. (2020). "Interpreting and Explaining Deep Neural Networks." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Choi, S. J., Kim, K., and Park, S. (2020). "Is systemic risk systematic? Evidence from the U.S. stock markets." In: International Journal of Finance & Economics, Early View.

Chollet, F. (2017). Deep Learning with Python. Manning Publications. 384 pp.

Choo, J. and Liu, S. (2018). "Visual analytics for explainable deep learning." In: IEEE Computer Graphics and Applications 38(4), pp. 84–92.

Chordia, T., Goyal, A., and Saretto, A. (2020). "Anomalies and false rejections." In: The Review of Financial Studies 33(5), pp. 2134–2179.

Clemente, G. P., Grassi, R., and Hitaj, A. (2018). "Asset allocation: new evidence through network approaches." In: arXiv e-Print.

Clemente, G. P., Grassi, R., and Hitaj, A. (2019). "Smart network based portfolios." In: arXiv e-Print.

Clemente, G. P., Grassi, R., and Hitaj, A. (2021). "Asset allocation: new evidence through network approaches." In: Annals of Operations Research, Early View.

Cohen, L., Malloy, C. J., and Nguyen, Q. (2019). "Lazy Prices." In: SSRN e-Print.

Cohort, P., Corbetta, J., and Laachir, I. (2020). "Analytical scores for stress scenarios." In: arXiv e-Print.

Colas, C., Fournier, P., Sigaud, O., Chetouani, M., and Oudeyer, P.-Y. (2019). "CURIOUS: Intrinsically Motivated Modular Multi-Goal Reinforcement Learning." In: arXiv e-Print.

Collin-Dufresne, P., Daniel, K., and Saglam, M. (2020). "Liquidity regimes and optimal dynamic asset allocation." In: Journal of Financial Economics 136(2), pp. 379–406.

Collot, S. and Hemauer, T. (2021). "A literature review of new methods in empirical asset pricing: omitted-variable and errors-in-variable bias." In: Financial Markets and Portfolio Management, Early View.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2020). "AlphaPortfolio for Investment and Economically Interpretable AI." In: SSRN e-Print.

Cont, R. (2001). "Empirical properties of asset returns: stylized facts and statistical issues." In: Quantitative Finance 1(2), pp. 223–236.

Cont, R., Potters, M., and Bouchaud, J.-P. (1997). "Scaling in stock market data: stable laws and beyond." In: arXiv e-Print.

Coqueret, G. (2020a). "Persistence in Factor-Based Supervised Learning Models." In: SSRN e-Print.

Coqueret, G. (2020b). "Stock-specific sentiment and return predictability." In: Quantitative Finance 20(9), pp. 1531–1551.

Coqueret, G. and Guida, T. (2020). Machine Learning for Factor Investing: R Version. Chapman and Hall/CRC. 341 pp.

Cornell, B. (2020). "ESG Investing: Conceptual Issues." In: SSRN e-Print.

Cornell, B. and Damodaran, A. (2020). "Valuing ESG: Doing Good or Sounding Good?" In: The Journal of Impact and ESG Investing 1(1), pp. 76–93.

Costa, G. and Kwon, R. H. (2020). "A regime-switching factor model for mean-variance optimization." In: Journal of Risk 22(4), pp. 31–59.

Costa, Y. M. G., Bertolini, D., Britto, A. S., Cavalcanti, G. D. C., and Oliveira, L. E. S. (2020). "The dissimilarity approach: a review." In: Artificial Intelligence Review 53(4), pp. 2783–2808.

Coulombe, C. (2018). "Text Data Augmentation Made Simple By Leveraging NLP Cloud APIs." In: arXiv e-Print.

Craik, D. (2019). How AI Can Help Find ESG Opportunities. Tech. rep. Morningstar.

Cramer, S. and Trimborn, T. (2019). "Stylized Facts and Agent-Based Modeling." In: arXiv e-Print.

Creamer, G. G. (2015). "Can a corporate network and news sentiment improve portfolio optimization using the Black-Litterman model?" In: Quantitative Finance 15(8), pp. 1405–1416.

Cui, P., Shen, Z., Li, S., Yao, L., Li, Y., Chu, Z., and Gao, J. (2020). "Causal Inference Meets Machine Learning." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Czasonis, M., Kritzman, M., Pamir, B., and Turkington, D. (2020). "Enhanced Scenario Analysis." In: The Journal of Portfolio Management 46(4), pp. 69–79.

D'Acunto, F. and Rossi, A. G. (2021). "New Frontiers of Robo-Advising: Consumption, Saving, Debt Management, and Taxes." In: SSRN e-Print.

da Costa, A. A. B. and Dasgupta, P. (2020). "Flexible Learning of Temporal Causal Sequence Relationships from Time-Series." In: arXiv e-Print.

Da Silva, B. and Shi, S. S. (2019). "Towards Improved Generalization in Financial Markets with Synthetic Data Generation." In: arXiv e-Print.

Dahlke, J., Bogner, K., Mueller, M., Berger, T., Pyka, A., and Ebersberger, B. (2020). "Is the Juice Worth the Squeeze? Machine Learning (ML) In and For Agent-Based Modelling (ABM)." In: arXiv e-Print.

Dal Pra, G., Guidolin, M., Pedio, M., and Vasile, F. (2018). "Regime Shifts in Excess Stock Return Predictability: An Out-of-Sample Portfolio Analysis." In: The Journal of Portfolio Management 44(3), pp. 10–24.

Danilevsky, M., Qian, K., Aharonov, R., Katsis, Y., Kawas, B., and Sen, P. (2020). "A Survey of the State of Explainable AI for Natural Language Processing." In: arXiv e-Print.

Dantas, T. M. and Oliveira, F. L. C. (2018). "Improving time series forecasting: An approach combining bootstrap aggregation, clusters and exponential smoothing." In: International Journal of Forecasting 34(4), pp. 748–761.

Dapena, J. P., Serur, J. A., and Siri, J. R. (2019). "Risk On-Risk Off: A Regime Switching Model for Active Portfolio Management." In: SSRN e-Print.

Das, S. and Varma, S. (2020). “Dynamic Goals-Based Wealth Management using Reinforcement Learning.” In: Journal of Investment Management 18 (2), pp. 1–20.

Das, S. R., Kim, S., and Ostrov, D. N. (2019a). “Dynamic Systemic Risk: Networks in Data Science.” In: The Journal of Financial Data Science 1 (1), pp. 141–158.

Das, S. R., Mitchener, K. J., and Vossmeyer, A. (2019b). “Systemic Risk and the Great Depression.” In: SSRN e-Print.

Das, S., Islam, M. R., Jayakodi, N. K., and Doppa, J. R. (2019c). “Active Anomaly Detection via Ensembles: Insights, Algorithms, and Interpretability.” In: arXiv e-Print.

Davis, M. H. A. and Lleo, S. (2016). “A Simple Procedure for Combining Expert Opinion with Statistical Estimates to Achieve Superior Portfolio Performance.” In: The Journal of Portfolio Management 42(4), pp. 49–58.

De Cristofaro, E. (2020). “An Overview of Privacy in Machine Learning.” In: arXiv e-Print. arXiv: 2005.08679v1.

de Franco, C. (2019a). “ESG Controversies and Their Impact on Performance.” In: The Journal of Investing 29(2), pp. 33–45.

de Franco, C. (2019b). Machine Learning meets ESG: How computing power can extract alpha from complex ESG data. Tech. rep. Natixis Investment Managers.

De Franco, C., Geissler, C., Margot, V., and Monnier, B. (2020). “ESG investments: Filtering versus machine learning approaches.” In: arXiv e-Print.

De Meer Pardo, F. and Lopez, R. C. (2020). “Mitigating Overfitting on Financial Datasets with Generative Adversarial Networks.” In: The Journal of Financial Data Science 2 (1), pp. 76–85.

Dees, B. S., Stankovic, L., Constantinides, A. G., and Mandic, D. P. (2019). “Portfolio Cuts: A Graph-Theoretic Framework to Diversification.” In: arXiv e-Print.

Deguest, R., Martellini, L., Milhau, V., Suri, A., and Wang, H. (2015). Introducing a Comprehensive Investment Framework for Goals-Based Wealth Management. Tech. rep. EDHEC Risk Institute.

del Barrio, E., Gordaliza, P., and Loubes, J.-M. (2020). "Review of Mathematical frameworks for Fairness in Machine Learning." In: arXiv e-Print.

Demirer, R., Demos, G., Gupta, R., and Sornette, D. (2019). "On the predictability of stock market bubbles: evidence from LPPLS confidence multi-scale indicators." In: Quantitative Finance 19 (5), pp. 843–858.

Demos, G. and Sornette, D. (2017). "Birth or burst of financial bubbles: which one is easier to diagnose?" In: Quantitative Finance 17(5), pp. 657–675.

Dempster, M. A. H., Kloppers, D., Medova, E., Osmolovsky, I., and Ustinov, P. (2016). "Lifecycle Goal Achievement or Portfolio Volatility Reduction?" In: The Journal of Portfolio Management 42(2), pp. 99–117.

Denev, A. and Mutnikas, Y. (2016a). "A formalized, integrated and visual approach to stress testing." In: Risk Management 18, pp. 189–216.

Denev, A. and Angelini, O. (2016). "Underdetermination and Variability of the Results in Macro-to-Micro Stress Tests - A Machine Learning Approach." In: SSRN e-Print.

Denev, A. and Mutnikas, Y. (2016b). "A Formalized, Integrated and Visual Approach to Stress Testing." In: SSRN e-Print.

Denev, A., Papaioannou, A., and Angelini, O. (2018). "A Probabilistic Graphical Models Approach to Model Interconnectedness." In: SSRN e-Print.

Dichtl, H., Drobetz, W., Lohre, H., and Rother, C. (2020). "Active factor completion strategies." In: SSRN e-Print.

Dimson, E., Marsh, P., and Staunton, M. (2020). "Divergent ESG Ratings." In: The Journal of Portfolio Management 47(1), pp. 75–87.

DiPietro, D. M. (2019). "Alpha Cloning: Using Quantitative Techniques and SEC 13f Data for Equity Portfolio Optimization and Generation." In: The Journal of Financial Data Science 1(4), pp. 159–171.

Dixon, M. and Polson, N. (2020). "Deep Fundamental Factor Models." In: SIAM Journal on Financial Mathematics 11(3), SC–26–SC–37.

Dixon, M. F., Halperin, I., and Bilokon, P. (2020). Machine Learning in Finance: from theory to practice. Springer International Publishing. 548 pp.

Dixon, M. F. and Halperin, I. (2019). "The Four Horsemen of Machine Learning in Finance." In: SSRN e-Print.

Dixon, M. F. and Halperin, I. (2020). "G-Learner and GIRL: Goal Based Wealth Management with Reinforcement Learning." In: SSRN e-Print.

Dong, L., Gao, G., Li, Y., and Wen, Y. (2019). "Baconian: A Unified Open source Framework for Model-Based Reinforcement Learning." In: arXiv e-Print.

Dose, C. and Cincotti, S. (2005). "Clustering of financial time series with application to index and enhanced index tracking portfolio." In: Physica A: Statistical Mechanics and its Applications 355(1), pp. 145–151.

Doshi-Velez, F. and Kim, B. (2017). "Towards A Rigorous Science of Interpretable Machine Learning." In: arXiv e-Print.

Dosilovic, F. K., Brcic, M., and Hlupic, N. (2018). "Explainable artificial intelligence: A survey." In: 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, pp. 0210–0215.

Downey, S. (2020). "Gazing into the Future: Using Ensemble Techniques to Forecast Company Fundamentals." In: SSRN e-Print.

Draminski, M. and Koronacki, J. (2018). "rmcfs : an R package for Monte Carlo feature selection and interdependency discovery." In: Journal of Statistical Software 85(12).

Drobetz, W. and Otto, T. (2020). "Empirical Asset Pricing via Machine Learning: Evidence from the European Stock Market." In: SSRN e-Print.

Du, J., Jin, M., Kolm, P. N., Ritter, G., Wang, Y., and Zhang, B. (2020). "Deep Reinforcement Learning for Option Replication and Hedging." In: The Journal of Financial Data Science 22(44), pp. 44–57.

Du, M., Liu, N., and Hu, X. (2018). "Techniques for Interpretable Machine Learning." In: arXiv e-Print.

Duanmu, J., Li, Y., and Malakhov, A. (2020). "Active factor investing: Hedge funds versus the rest of us." In: Review of Financial Economics, Early View.

Duarte, F. G. and De Castro, L. N. (2020). "A Framework to Perform Asset Allocation Based on Partitional Clustering." In: IEEE Access: Practical Iinnovations, Open Solutions 8, pp. 110775–110788.

Dunkler, D., Plischke, M., Leffondre, K., and Heinze, G. (2014). "Augmented backward elimination: a pragmatic and purposeful way to develop statistical models." In: PLOS One 9(11), e113677.

Durante, F., Foscolo, E., Pappada, R., and Wang, H. (2015). A portfolio diversification strategy via tail dependence measures. Tech. rep. DEAMS Research Paper Series.

Durante, F., Pappada, R., and Torelli, N. (2013). "Clustering of financial time series in risky scenarios." In: Advances in Data Analysis and Classification 8(4), pp. 359–376.

Dwork, C., Ilvento, C., Rothblum, G. N., and Sur, P. (2020). "Abstracting Fairness: Oracles, Metrics, and Interpretability." In: arXiv e-Print. arXiv: `2004.01840 [cs.LG]`.

Echihabi, K., Zoumpatianos, K., Palpanas, T., and Benbrahim, H. (2018). "The lernaean hydra of data series similarity search." In: Proceedings of the VLDB Endowment 12(2), pp. 112–127.

Edelmann, D., Fokianos, K., and Pitsillou, M. (2019). "An Updated Literature Review of Distance Correlation and Its Applications to Time Series." In: International Statistical Review 87(2), pp. 237–262.

Efron, B. and Hastie, T. (2016). Computer Age Statistical Inference: Algorithms, Evidence and Data Science. Cambridge University Press. 496 pp.

Eisenstein, J. (2019). Introduction to Natural Language Processing. MIT Press. 536 pp.

Ekster, G. and Kolm, P. N. (2020). "Alternative Data in Investment Management: Usage, Challenges and Valuation." In: SSRN e-Print.

Elagamy, M. N., Stanier, C., and Sharp, B. (2018). "Stock market random forest-text mining system mining critical indicators of stock market movements." In: 2nd International Conference on Natural Language and Speech Processing (ICNLSP). IEEE, pp. 1–8.

Elliott, G. (2017). "Forecast combination when outcomes are difficult to predict." In: Empirical Economics 53(1), pp. 7–20.

Engle, R. F. (2020). "Stress Testing with Market Data." In: SSRN e-Print.

Engle, R. F. and Roussellet, G. (2016). "Scenario Generation for Long-Run Interest Rate Risk Assessment." In: SSRN e-Print.

Eom, C. and Park, J. W. (2017). "Effects of common factors on stock correlation networks and portfolio diversification." In: International Review of Financial Analysis 49, pp. 1–11.

Fabozzi, F. J. and Lopez de Prado, M. (2018). "Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests." In: The Journal of Portfolio Management 45(1), pp. 141–147.

Facchinato, S. and Pola, G. (2014). "Managing uncertainty with diversification across macroeconomic scenarios (DAMS): from asset segmentation to portfolio." In: SSRN e-Print.

Fama, E. F. and French, K. R. (2020). "Comparing Cross-Section and Time-Series Factor Models." In: The Review of Financial Studies 33(5), pp. 1891–1926.

Fan, J., Xue, L., and Yao, J. (2017). "Sufficient forecasting using factor models." In: Journal of Econometrics 201(2), pp. 292–306.

Fang, J. and Lin, J. (2020). "Prior knowledge distillation based on financial time series." In: arXiv e-Print.

Fang, M., Kuo, L., Shih, F., and Taylor, S. (2020). "Sector categorization using gradient boosted trees trained on fundamental firm data." In: Algorithmic Finance, Early View.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2018a). "Data augmentation using synthetic data for time series classification with deep residual networks." In: arXiv e-Print.

Fawaz, H. I., Forestier, G., Weber, J., Idoumghar, L., and Muller, P.-A. (2018b). "Transfer learning for time series classification." In: IEEE International Conference on Big Data (Big Data). IEEE, pp. 1367–1376.

Feng, G., Giglio, S., and Xiu, D. (2017). "Taming the Factor Zoo." In: SSRN e-Print.

Feng, G., Giglio, S., and Xiu, D. (2020). "Taming the factor zoo: A test of new factors." In: The Journal of Finance 75(3), pp. 1327–1370.

Feng, G., Polson, N. G., and Xu, J. (2018). "Deep Factor Alpha." In: arXiv e-Print.

Ferrario, A. and Naegelin, M. (2020). "The Art of Natural Language Processing: Classical, Modern and Contemporary Approaches to Text Document Classification." In: SSRN e-Print.

Feuerriegel, S. and Gordon, J. (2018). "Long-term stock index forecasting based on text mining of regulatory disclosures." In: Decision Support Systems 112, pp. 88–97.

Feurer, M. and Hutter, F. (2019). "Hyperparameter Optimization." In: Automated machine learning: methods, systems, challenges. Ed. by F. Hutter, L. Kotthoff, and J. Vanschoren. Springer, pp. 3–33.

Fievet, L. and Sornette, D. (2018). "Calibrating emergent phenomena in stock markets with agent based models." In: PLOS ONE 13(3). Ed. by F. Wen, e0193290+.

Filimonov, V. and Sornette, D. (2013). "A stable and robust calibration scheme of the log-periodic power law model." In: Physica A: Statistical Mechanics and its Applications 392(17), pp. 3698–3707.

Fischer, T. and Krauss, C. (2017). "Deep learning with long short-term memory networks for financial market predictions." In: European Journal of Operational Research 270 (2), pp. 654–669.

Fischer, T. G. (2018). Reinforcement learning in financial markets - a survey. Tech. rep. Friedrich-Alexander University.

Fleiss, A., Cui, H., Stoikov, S., and DiPietro, D. M. (2020). "Constructing Equity Portfolios from SEC 13F Data Using Feature Extraction and Machine Learning." In: The Journal of Financial Data Science 2 (1), pp. 45–60.

Fletcher, J. (2019a). "How many factors are important in U.K. stock returns?" In: The European Journal of Finance 25(13), pp. 1234–1249.

Fletcher, J. (2019b). "Model comparison tests of linear factor models in U.K. stock returns." In: Finance Research Letters 28, pp. 281–291.

Flint, E. J. and Mare, E. (2019). "Regime-Based Tactical Allocation for Equity Factors and Balanced Portfolios." In: South African Actuarial Journal 19(1), pp. 27–52.

Flood, M. D. and Korenko, G. G. (2015). "Systematic scenario selection: stress testing and the nature of uncertainty." In: Quantitative Finance 15(1), pp. 43–59.

Fons, E., Dawson, P., Yau, J., Zeng, X.-j., and Keane, J. (2021). "A novel dynamic asset allocation system using Feature Saliency Hidden Markov models for smart beta investing." In: Expert Systems with Applications 163, pp. 113720+.

Foorthuis, R. (2020). "On the Nature and Types of Anomalies: A Review." In: arXiv e-Print.

Fornaro, P. and Luomaranta, H. (2020). "Nowcasting Finnish real economic activity: a machine learning approach." In: Empirical Economics 58(1), pp. 55–71.

Forsberg, D., Gallagher, D. R., and Warren, G. (2018). "Relative Hedge Fund Skill and the Informativeness of Cohort Alpha." In: SSRN e-Print.

Fortunato, S. and Hric, D. (2016). "Community detection in networks: A user guide." In: Physics Reports 659, pp. 1–44.

Fox, G. and Jha, S. (2019). "Learning Everywhere: A Taxonomy for the Integration of Machine Learning and Simulations." In: 15th International Conference on eScience (eScience). IEEE.

Franco-Pedroso, J., Gonzalez-Rodriguez, J., Planas, M., Cubero, J., Cobo, R., and Pablos, F. (2019a). "Generating Virtual Scenarios of Multivariate Financial Data for Quantitative Trading Applications." In: The Journal of Financial Data Science 1 (2), pp. 55–77.

Franco-Pedroso, J., Gonzalez-Rodriguez, J., Planas, M., Cubero, J., Cobo, R., and Pablos, F. (2019b). "The ETS challenges: a machine learning approach to the evaluation of simulated financial time series for improving generation processes." In: The Journal of Financial Data Science 1(3) (3), pp. 68–86.

Francois-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., and Pineau, J. (2018). "An introduction to deep reinforcement learning." In: Foundations and Trends in Machine Learning 11(3-4), pp. 219–354.

Freyberger, J., Neuhierl, A., and Weber, M. (2018). "Dissecting Characteristics Nonparametrically." In: SSRN e-Print.

Freyberger, J., Neuhierl, A., and Weber, M. (2020). "Dissecting Characteristics Nonparametrically." In: The Review of Financial Studies 33(5), pp. 2326–2377.

Friedler, S. A., Roy, C. D., Scheidegger, C., and Slack, D. (2019). "Assessing the Local Interpretability of Machine Learning Models." In: arXiv e-Print.

Frydman, R., Mangee, N., and Stillwagon, J. (2021). "How Market Sentiment Drives Forecasts of Stock Returns." In: Journal of Behavioral Finance (In Press), Early View.

Gade, K., Geyik, S. C., Kenthapadi, K., Mithal, V., and Taly, A. (2019). "Explainable AI in industry." In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '19. ACM Press, pp. 3203–3204.

Gao, G., Mishra, B., and Ramazzotti, D. (2018). "Causal data science for financial stress testing." In: Journal of Computational Science 26, pp. 294–304.

Gao, H., He, J., and Chen, K. (2020). "Exploring Machine Learning Techniques for Text-Based Industry Classification." In: SSRN e-Print.

Garvey, G. and Madhavan, A. (2019). "Reconstructing Emerging and Developed Markets Using Hierarchical Clustering." In: The Journal of Financial Data Science 1 (4), pp. 84–102.

Gasthaus, J., Benidis, K., Wang, Y., Rangapuram, S. S., Salinas, D., Flunkert, V., and Januschowski, T. (2019). "Probabilistic Forecasting with Spline Quantile Function RNNs." In: Proceedings of Machine Learning Research 89, pp. 1901–1910.

Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A., and Veeramachaneni, K. (2020). "TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks." In: arXiv e-Print.

Ghorbani, A., Wexler, J., and Kim, B. (2019). "Automating Interpretability: Discovering and Testing Visual Concepts Learned by Neural Networks." In: arXiv e-Print.

Gibbs, C. G. (2017). "Conditionally Optimal Weights and Forward-Looking Approaches to Combining Forecasts." In: SSRN e-Print.

Gidea, M. and Katz, Y. (2018). "Topological data analysis of financial time series: Landscapes of crashes." In: Physica A: Statistical Mechanics and its Applications 491, pp. 820–834.

Gidwani, B. (2020). "Some Issues with Using ESG Ratings in an Investment Process." In: The Journal of Investing 29(6), pp. 76–84.

Giglio, S., Liao, Y., and Xiu, D. (2021). "Thousands of Alpha Tests." In: The Review of Financial Studies, Early View.

Giglio, S. W. and Xiu, D. (2019). "Asset Pricing with Omitted Factors." In: SSRN e-Print.

Gilliland, M. (2020). "The value added by machine learning approaches in forecasting." In: International Journal of Forecasting 36(1), pp. 161–166.

Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., and Kagal, L. (2018). "Explaining Explanations: An Approach to Evaluating Interpretability of Machine Learning." In: arXiv e-Print.

Giudici, P., Polinesi, G., and Spelta, A. (2020). "Network models to improve robot advisory portfolio asset allocation." In: SSRN e-Print.

Giuzio, M., Eichhorn-Schott, K., Paterlini, S., and Weber, V. (2018). "Tracking hedge funds returns using sparse clones." In: Annals of Operations Research 266, pp. 349–371.

Glasserman, P., Kang, C., and Kang, W. (2015). "Stress scenario selection by empirical likelihood." In: Quantitative Finance 15(1), pp. 25–41.

Gnecco, N., Meinshausen, N., Peters, J., and Engelke, S. (2020). "Causal discovery in heavy-tailed models." In: arXiv e-Print.

Goan, E. and Fookes, C. (2020). “Bayesian Neural Networks: An Introduction and Survey.” In: arXiv e-Print.

Gokhale, S., Korukonda, A., Li, J., and Zhu, W. (2019). “Factor and Industry Allocation using Markov-Switching Model.” In: SSRN e-Print.

Golub, B., Greenberg, D., and Ratcliffe, R. (2018). “Market-Driven Scenarios: An Approach for Plausible Scenario Construction.” In: The Journal of Portfolio Management 44(5), pp. 6–20.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep Learning. Adaptive Computation and Machine Learning series. The MIT Press. 800 pp.

Gosling, S. (2010). “A Scenarios Approach to Asset Allocation.” In: The Journal of Portfolio Management 37(1), pp. 53–66.

Gospodinov, N. and Robotti, C. (2021). “Common pricing across asset classes: Empirical evidence revisited.” In: Journal of Financial Economics 132(2), Forthcoming.

Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S. F. M., and Kurths, J. (2018). “Abrupt transitions in time series with uncertainties.” In: Nature Communications 9(1) (48).

Gotze, T., Gurtler, M., and Witowski, E. (2020). “Improving CAT bond pricing models via machine learning.” In: Journal of Asset Management 21, pp. 428–446.

Grammig, J., Hanenberg, C., Schlag, C., and Sonksen, J. (2020). “Diverging Roads: Theory-Based vs. Machine Learning-Implied Stock Risk Premia.” In: SSRN e-Print.

Grealish, A. and Kolm, P. N. (2021). “Robo-Advisory: From Investing Principles and Algorithms to Future Developments.” In: SSRN e-Print.

Greco, C., Polonioli, A., and Tagliabue, J. (2019). “Less (Data) Is More: Why Small Data Holds the Key to the Future of Artificial Intelligence.” In: arXiv e-Print.

Green, J., Hand, J. R. M., and Zhang, X. F. (2017). “The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns.” In: The Review of Financial Studies 30(12), pp. 4389–4436.

Grundke, P. (2012). "Further recipes for quantitative reverse stress testing." In: Journal of Risk Model Validation 6(2).

Gu, S., Kelly, B., and Xiu, D. (2021). "Autoencoder asset pricing models." In: Journal of Econometrics, Early View.

Gu, S., Kelly, B. T., and Xiu, D. (2020). "Empirical asset pricing via machine learning." In: The Review of Financial Studies 33 (5), pp. 2223–2273.

Guida, T. (2019). Big Data and Machine Learning in Quantitative Investment. Wiley. 285 pp.

Guo, R., Cheng, L., Li, J., Hahn, P. R., and Liu, H. (2020a). "A Survey of Learning Causality with Data: Problems and Methods." In: arXiv e-Print.

Guo, T., Jamet, N., Betrix, V., Piquet, L.-A., and Hauptmann, E. (2020b). "ESG2Risk: A Deep Learning Framework from ESG News to Stock Volatility Prediction." In: arXiv e-Print.

Gupta, Y. (2018). "Selection of important features and predicting wine quality using machine learning techniques." In: Procedia Computer Science 125, pp. 305–312.

Haldane, A. G. and Turrell, A. E. (2018). "Drawing on different disciplines: macroeconomic agent-based models." In: Journal of Evolutionary Economics 29(1), pp. 1–28.

Hall, P. (2020). Awesome machine learning interpretability. URL: https://github.com/jphall663/awesome-machine-learning-interpretability.

Hall, P., Gill, N., and Meng, L. (2018). "Testing machine learning interpretability techniques." In: O'Reilly.

Halvari, T., Nurminen, J. K., and Mikkonen, T. (2020). "Testing the Robustness of AutoML Systems." In: arXiv e-Print.

Hamilton, J. D. (2016). "Macroeconomic Regimes and Regime Shifts." In: Handbook of Macroeconomics. Elsevier, pp. 163–201.

Hammerschmid, R. and Lohre, H. (2018). "Regime Shifts and Stock Return Predictability." In: International Review of Economics and Finance 56, pp. 138–160.

Hansen, P. R., Lunde, A., and Nason, J. M. (2011). "The Model Confidence Set." In: Econometrica 79(2), pp. 453–497.

Hanussek, M., Blohm, M., and Kintz, M. (2020). "Can AutoML outperform humans? An evaluation on popular OpenML datasets using AutoML Benchmark." In: arXiv e-Print.

Harper, H. (2020). "One Institutional Investor Approach to Integrating ESG in the Investment Process." In: The Journal of Portfolio Management 46(4), pp. 110–123.

Harvey, C. R., Hoyle, E., Rattray, S., and van Hemert, O. (2020a). "Strategic Risk Management: Out-of-Sample Evidence from the COVID-19 Equity Selloff." In: SSRN e-Print.

Harvey, C. R. and Liu, Y. (2015). "Rethinking Performance Evaluation." In: SSRN e-Print.

Harvey, C. R. and Liu, Y. (2018). "Detecting Repeatable Performance." In: The Review of Financial Studies 31 (7), pp. 2499–2552.

Harvey, C. R. and Liu, Y. (2019). "Cross-sectional alpha dispersion and performance evaluation." In: Journal of Financial Economics 134(2), pp. 273–296.

Harvey, C. R. and Liu, Y. (2020a). "A census of the factor zoo." In: SSRN e-Print.

Harvey, C. R. and Liu, Y. (2020b). "False (and missed) discoveries in financial economics." In: The Journal of Finance 75(5), pp. 2503–2553.

Harvey, C. R. and Liu, Y. (2020c). "Lucky Factors." In: SSRN e-Print.

Harvey, C. R., Liu, Y., and Saretto, A. (2020b). "An Evaluation of Alternative Multiple Testing Methods for Finance Applications." In: The Review of Asset Pricing Studies 10(2), pp. 199–248.

Harvey, C. R., Liu, Y., and Zhu, H. (2016). "... and the Cross-Section of Expected Returns." In: The Review of Financial Studies 29 (1), pp. 5–68.

Hastie, T., Tibshirani, R., and Friedman, J. (2011). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition. Springer. 552 pp.

Hatcher, W. G. and Yu, W. (2018). "A survey of deep learning: platforms, applications and emerging research trends." In: IEEE Access: Practical Iinnovations, Open Solutions 6, pp. 24411–24432.

Hazard, C. J., Fusting, C., Resnick, M., Auerbach, M., Meehan, M., and Korobov, V. (2019). "Natively Interpretable Machine Learning and Artificial Intelligence: Preliminary Results and Future Directions." In: arXiv e-Print.

Hazarika, S., Biswas, A., Wolfram, P. J., Lawrence, E., and Urban, N. (2020). "Relationship-aware Multivariate Sampling Strategy for Scientific Simulation Data." In: arXiv e-Print.

He, Q.-Q., Pang, P. C.-I., and Si, Y.-W. (2019a). "Transfer learning for financial time series forecasting." In: PRICAI 2019: trends in artificial intelligence: 16th Pacific Rim international conference on artificial intelligence. Ed. by A. C. Nayak and A. Sharma. Vol. 11671. Springer, pp. 24–36.

He, Z., Xie, L., Chen, X., Zhang, Y., Wang, Y., and Tian, Q. (2019b). "Data Augmentation Revisited: Rethinking the Distribution Gap between Clean and Augmented Data." In: arXiv e-Print.

Heaton, J. B. (2018). "Quantitative Investing and the Limits of (Deep) Learning from Financial Data." In: The Journal of Financial Transformation 47.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2017). "Why Indexing Works." In: arXiv e-Print.

Hellstrom, T., Dignum, V., and Bensch, S. (2020). "Bias in Machine Learning What is it Good (and Bad) for?" In: arXiv e-Print. arXiv: 2004.00686.

Hens, T., Schenk-Hoppe, K. R., and Woesthoff, M.-H. (2020). "Escaping the backtesting illusion." In: The Journal of Portfolio Management 46(4), pp. 81–93.

Hernandez-Garcia, A. and Konig, P. (2018). "Data augmentation instead of explicit regularization." In: arXiv e-Print.

Heuillet, A., Couthouis, F., and Diaz-Rodriguez, N. (2020). "Explainability in Deep Reinforcement Learning." In: arXiv e-Print.

Hewamalage, H., Bergmeir, C., and Bandara, K. (2019). "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions." In: arXiv e-Print.

Higham, C. F. and Higham, D. J. (2018). "Deep Learning: An Introduction for Applied Mathematicians." In: arXiv e-Print.

Hilliard, J., Narayanasamy, A., and Zhang, S. (2020). "The role of market sentiment in asset allocations and stock returns." In: Journal of Behavioral Finance 21(4), pp. 423–441.

Hoepner, A. G. F., McMillan, D., Vivian, A., and Simen, C. W. (2021). "Significance, relevance and explainability in the machine learning age: an econometrics and financial data science perspective." In: The European Journal of Finance 27(1-2), Early View1–7.

Hoffmann, J., Bar-Sinai, Y., Lee, L., Andrejevic, J., Mishra, S., Rubinstein, S. M., and Rycroft, C. H. (2018). "Machine Learning in a data-limited regime: Augmenting experiments with synthetic data uncovers order in crumpled sheets." In: arXiv e-Print.

Horel, E. and Giesecke, K. (2018). "Towards explainable AI: significance tests for neural networks." In: SSRN e-Print.

Horel, E. and Giesecke, K. (2019). "Computationally Efficient Feature Significance and Importance for Machine Learning Models." In: arXiv e-Print.

Horel, E. and Giesecke, K. (2020). "Significance Tests for Neural Networks." In: arXiv e-Print.

Horel, E., Giesecke, K., Storchan, V., and Chittar, N. (2019). "Explainable Clustering and Application to Wealth Management Compliance." In: arXiv e-Print.

Hou, K., Mo, H., Xue, C., and Zhang, L. (2021). "An Augmented q-Factor Model with Expected Growth." In: Review of Finance, Early View.

Hou, X., Wang, K., Zhang, J., and Wei, Z. (2020). "An Enriched Time-Series Forecasting Framework for Long-Short Portfolio Strategy." In: IEEE Access: Practical Iinnovations, Open Solutions 8, pp. 31992–32002.

Houlihan, P. and Creamer, G. G. (2017). "Risk premium of social media sentiment." In: The Journal of Investing 26(3), pp. 21–28.

Hsiao, C. and Wan, S. K. (2014). "Is there an optimal forecast combination?" In: Journal of Econometrics 178, pp. 294–309.

Hu, Y.-J. and Lin, S.-J. (2019). “Deep reinforcement learning for optimizing finance portfolio management.” In: Amity International Conference on Artificial Intelligence (AICAI). IEEE, pp. 14–20.

Huang, B., Zhang, K., Zhang, J., Ramsey, J., Sanchez-Romero, R., Glymour, C., and Scholkopf, B. (2020a). “Causal Discovery from Heterogeneous/Nonstationary Data.” In: Journal of Machine Learning Research 21(89), pp. 1–53.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2017). “Forecasting Stock Returns in Good and Bad Times: The Role of Market States.” In: SSRN e-Print.

Huang, D., Lehkonen, H., Pukthuanthong, K., and Zhou, G. (2018). “Sentiment across asset markets.” In: SSRN e-Print.

Huang, Y., Fu, Z., and Franzke, C. L. E. (2020b). “Detecting causality from time series in a machine learning framework.” In: Chaos: An Interdisciplinary Journal of Nonlinear Science 30(6), p. 063116.

Husmann, S., Shivarova, A., and Steinert, R. (2020). “Company classification using machine learning.” In: arXiv e-Print.

Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). Automated machine learning: methods, systems, challenges. Springer International Publishing. 219 pp.

Hwang, S. and Rubesam, A. (2019). “Searching the Factor Zoo.” In: SSRN e-Print.

Hyndman, R. J. and Athanasopoulos, G. (2020). Forecasting: Principles and Practice (Third Edition). OTexts. 380 pp.

Ignatiev, A. (2020). “Towards Trustable Explainable AI.” In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization.

Imokoyende, M. (2019). “Variational Autoencoder In Finance.” In: Towards Data Science.

Ioannidis, J. P. A., Cripps, S., and Tanner, M. A. (2020). “Forecasting for COVID-19 has failed.” In: International Journal of Forecasting, Early View.

Iori, G. and Mantegna, R. N. (2018). "Empirical analyses of networks in finance." In: Handbook of Computational Economics. Vol. 4. Elsevier, pp. 637–685.

Irlam, G. (2020). "Machine learning for retirement planning." In: The Journal of Retirement 8(1), pp. 32–29.

Ish-Horowicz, J., Udwin, D., Flaxman, S., Filippi, S., and Crawford, L. (2019). "Interpreting Deep Neural Networks Through Variable Importance." In: arXiv e-Print.

Iwana, B. K. and Uchida, S. (2020). "An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks." In: arXiv e-Print.

Jabbar, A., Li, X., and Omar, B. (2020). "A Survey on Generative Adversarial Networks: Variants, Applications, and Training." In: arXiv e-Print.

Jacob, D., Lessmann, S., and Hardle, W. K. (2020a). "Causal Inference Using Machine Learning. An Evaluation of Recent Methods Through Simulations." In: SSRN e-Print.

Jacob, V., Song, F., Stiegler, A., Diao, Y., and Tatbul, N. (2020b). "AnomalyBench: An Open Benchmark for Explainable Anomaly Detection." In: arXiv e-Print.

Jaeger, M., Krugel, S., Marinelli, D., Papenbrock, J., and Schwendner, P. (2020). "Understanding machine learning for diversified portfolio construction by explainable AI." In: SSRN e-Print.

Jaeger, M., Krugel, S., Marinelli, D., Papenbrock, J., and Schwendner, P. (2021). "Interpretable Machine Learning for Diversified Portfolio Construction." In: SSRN e-Print.

Jain, P. and Jain, S. (2019). "Can Machine Learning-Based Portfolios Outperform Traditional Risk-Based Portfolios? The Need to Account for Covariance Misspecification." In: Risks 7(3), pp. 74+.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2015). An Introduction to Statistical Learning: with Applications in R. Springer. 426 pp.

Jansen, S. (2020). Machine Learning for Algorithmic Trading. Packt Publishing. 820 pp.

Janssen, R., Kramer, B., and Boender, G. (2013). "Life Cycle Investing: From Target-Date to Goal-Based Investing." In: The Journal of Wealth Management 16(1), pp. 23–32.

Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., and Callot, L. (2020). "Criteria for classifying forecasting methods." In: International Journal of Forecasting 36(1) (1), pp. 167–177.

Javed, A., Lee, B. S., and Rizzo, D. M. (2020). "A Benchmark Study on Time Series Clustering." In: arXiv e-Print.

Jeub, L. G. S., Balachandran, P., Porter, M. A., Mucha, P. J., and Mahoney, M. W. (2015). "Think locally, act locally: Detection of small, medium-sized, and large communities in large networks." In: Physical Review E 91(1) (012821).

Jha, V. (2017). "Machine Learning Algorithm - Backbone of emerging technologies." In: Techleer.

Jilla, K., Nayak, S. C., and Bathula, A. (2017). "Stylized Facts Of Financial Time Series: A Comprehensive Analysis." In: International Journal of Recent Trends in Engineering & Research, pp. 302–307.

Jomaa, H. S., Grabocka, J., and Schmidt-Thieme, L. (2019). "Hyp-RL : Hyperparameter Optimization by Reinforcement Learning." In: arXiv e-Print.

Joshi, S. (2019). "Time Series Analysis and Forecasting of the US Housing Starts using Econometric and Machine Learning Model." In: arXiv e-Print.

Jurczenko, E. and Teiletche, J. (2018). "Active Risk-Based Investing." In: The Journal of Portfolio Management 44(3), pp. 56–65.

Jurczenko et al. (2020). Machine Learning for Asset Management. Ed. by E. Jurczenko. Wiley. 445 pp.

Kadous, K., Mercer, M., and Zhou, Y. (2017). "Undue influence? the effect of social media advice on investment decisions." In: SSRN e-Print.

Kakushadze, Z. and Yu, W. (2017a). "Open Source Fundamental Industry Classification." In: MDPI Data 22 (2).

Kakushadze, Z. and Yu, W. (2017b). "Statistical Industry Classification." In: SSRN e-Print.

Kakushadze, Z. and Yu, W. (2020). "Machine learning treasury yields." In: SSRN e-Print.

Kamath, U., Liu, J., and Whitaker, J. (2019). Deep Learning for NLP and Speech Recognition. Springer. 613 pp.

Kang, Y., Hyndman, R. J., and Li, F. (2020). "GRATIS: GeneRAting TIme Series with diverse and controllable characteristics." In: Statistical Analysis and Data Mining: The ASA Data Science Journal 13(4), pp. 354–376.

Kang, Y., Spiliotis, E., Petropoulos, F., Athiniotis, N., Li, F., and Assimakopoulos, V. (2019). "Deja vu: forecasting with similarity." In: arXiv e-Print.

Karathanasopoulos, A., Mitra, S., Skindilias, K., and Lo, C. C. (2017). "Modelling and Trading the English and German Stock Markets with Novelty Optimization Techniques." In: Journal of Forecasting 36(8) (8), pp. 974–988.

Karolyi, A. and Van Nieuwerburgh, S. (2020). "New Methods for the Cross-Section of Returns." In: The Review of Financial Studies 33(5), pp. 1879–1890.

Katakami, S., Sakamoto, H., and Okada, M. (2019). "Bayesian Hyperparameter Estimation using Gaussian Process and Bayesian Optimization." In: Journal of the Physical Society of Japan 88(7), p. 074001.

Kauffmann, J., Esders, M., Montavon, G., Samek, W., and Muller, K.-R. (2019). "From Clustering to Cluster Explanations via Neural Networks." In: arXiv e-Print.

Kaut, M. (2020). "Scenario generation using historical data paths." In: Optimization Online e-Print.

Kaut, M. and Lium, A.-G. (2015). "Scenario generation with distribution functions and correlations." In: Kybernetika, pp. 1049–1064.

Kawamoto, T. and Kabashima, Y. (2017a). "Comparative analysis on the selection of number of clusters in community detection." In: arXiv e-Print.

Kawamoto, T. and Kabashima, Y. (2017b). "Cross-validation estimate of the number of clusters in a network." In: Scientific Reports 7(1).

Kaya, H. (2015). "Eccentricity in Asset Management." In: Journal of Network Theory in Finance 1(1), pp. 1–32.

Kaya, H. (2017). “Managing ambiguity in asset allocation.” In: Journal of Asset Management 18(3), pp. 163–187.

Kee, T. (2019). “Peer Firm Identification Using Word Embeddings.” In: 2019 IEEE International Conference on Big Data (Big Data). IEEE.

Kegel, L., Hahmann, M., and Lehner, W. (2017). “Generating What-If Scenarios for Time Series Data.” In: Proceedings of the 29th International Conference on Scientific and Statistical Database Management - ’17. New York, New York, USA: ACM Press, pp. 1–12.

Kellner, R. and Rosch, D. (2020). “A Bayesian Re-Interpretation of Empirical Financial Research.” In: Finance Research Letters, Early View.

Khaleghi, A., Ryabko, D., Mary, J., and Preux, P. (2016). “Consistent algorithms for clustering time series.” In: The Journal of Machine Learning Research 17, pp. 1–32.

Kilburn, F. (2019a). “Can nowcasting unlock factor timing?” In: Risk.

Kilburn, F. (2019b). “Quant funds look to AI to master correlations.” In: Risk.

Kilburn, F. (2020a). “Lighting up the black box: a must for investors?” In: Risk.

Kilburn, F. (2020b). “Ripping up the old asset class labels.” In: Risk.

Kim, B. and Doshi-Velez, F. (2017). “Interpretable Machine Learning: The fuss, the concrete and the questions.” In: Thirty-fourth International Conference on Machine Learning.

Kim, H. (2019). “Interpretable Models in Probabilistic Machine Learning.” PhD thesis. University of Oxford.

Kim, J.-Y. and Cho, S.-B. (2019). “Evolutionary optimization of hyperparameters in deep learning models.” In: IEEE Congress on Evolutionary Computation (CEC). IEEE, pp. 831–837.

Kim, W. C., Kwon, D.-G., Lee, Y., Kim, J. H., and Lin, C. (2020). “Personalized goal-based investing via multi-stage stochastic goal programming.” In: Quantitative Finance 20(3) (3), pp. 515–526.

Kingma, D. P. and Welling, M. (2019). “An introduction to variational autoencoders.” In: Foundations and Trends in Machine Learning 12(4), pp. 307–392.

Kinlaw, W. and Turkington, D. (2014). "Correlation surprise." In: Journal of Asset Management 14(6), pp. 385–399.

Klein, A., Dai, Z., Hutter, F., Lawrence, N., and Gonzalez, J. (2019). "Meta-Surrogate Benchmarking for Hyperparameter Optimization." In: arXiv e-Print.

Klevak, J., Livnat, J., and Suslava, K. (2019). "A practical approach to advanced text mining in finance." In: The Journal of Financial Data Science 1 (1), pp. 122–129.

Knight, W. (2019). "An AI Pioneer Wants His Algorithms to Understand the 'Why'." In: Wired.

Kolanovic, M. and Krishnamachari, R. T. (2017). Big Data and AI Strategies: Machine Learning and Alternative Data Approach to Investing. Tech. rep. JP Morgan.

Kolm, P. and Ritter, G. (2020a). "Factor investing with Black-Litterman-Bayes: incorporating factor views and priors in portfolio construction." In: SSRN e-Print.

Kolm, P. N. and Ritter, G. (2019). "Dynamic Replication and Hedging: A Reinforcement Learning Approach." In: The Journal of Financial Data Science 1 (1), pp. 159–171.

Kolm, P. N. and Ritter, G. (2020b). "Modern Perspectives on Reinforcement Learning in Finance." In: The Journal of Machine Learning in Finance 1(1).

Kolrep, M., Lohre, H., Radatz, E., and Rother, C. (2020). "Economic Versus Statistical Clustering in Multi-Asset Multi-Factor Strategies." In: Risk & Reward, pp. 26–32.

Kondratyev, A. and Schwarz, C. (2019). "The Market Generator." In: SSRN e-Print.

Kondratyev, A. and Schwarz, C. (2020). "The Market Generator." In: Risk.

Kondratyev, A., Schwarz, C., and Horvath, B. (2020). "Data Anonymisation, Outlier Detection and Fighting Overfitting with Restricted Boltzmann Machines." In: Risk.

Konstantinov, G., Chorus, A., and Rebmann, J. (2020). "A network and machine learning approach to factor, asset, and blended allocation." In: The Journal of Portfolio Management 46 (6), pp. 54–71.

Konstantinov, G. and Rusev, M. (2020). "The Bond-Equity-Fund Relation Using the Fama-French-Carhart Factors: A Practical Network Approach." In: The Journal of Financial Data Science 2 (1), pp. 24–44.

Konstantinov, G. S. and Simonian, J. (2020). "A Network Approach to Analyzing Hedge Fund Connectivity." In: The Journal of Financial Data Science 22(3) (3), pp. 55–72.

Koochali, A., Dengel, A., and Ahmed, S. (2020). "If You Like It, GAN It. Probabilistic Multivariate Times Series Forecast With GAN." In: arXiv e-Print.

Kopeliovich, Y., Novosyolov, A., Satchkov, D., and Schachter, B. (2015). "Robust Risk Estimation and Hedging: A Reverse Stress Testing Approach." In: The Journal of Derivatives 22(4), pp. 10–25.

Koshiyama, A. and Firoozye, N. (2019). "Avoiding Backtesting Overfitting by Covariance-Penalties: an empirical investigation of the ordinary and total least squares cases." In: The Journal of Financial Data Science 1(4), pp. 63–83.

Koshiyama, A., Firoozye, N., and Treleaven, P. (2019). "Generative Adversarial Networks for Financial Trading Strategies Fine-Tuning and Combination." In: arXiv e-Print.

Koshiyama, A., Firoozye, N., and Treleaven, P. (2020). "Algorithms in Future Capital Markets." In: SSRN e-Print.

Krishna, R. and Menzies, T. (2018). "Bellwethers: A baseline method for transfer learning." In: IEEE Transactions on Software Engineering 45 (11), pp. 1081–1105.

Kristoufek, L. (2018). "Power-law cross-correlations: Issues, solutions and future challenges." In: arXiv e-Print.

Kritzman, M., Page, S., and Turkington, D. (2012). "Regime Shifts: Implications for Dynamic Strategies." In: Financial Analysts Journal 68(3).

Krohn, J., Beyleveld, G., and Bassens, A. (2019). Deep Learning Illustrated: A Visual, Interactive Guide to Artificial Intelligence. Addison-Wesley. 416 pp.

Kuchnik, M. and Smith, V. (2018). "Efficient Augmentation via Data Subsampling." In: arXiv e-Print.

Lacasa, L., Nicosia, V., and Latora, V. (2015). "Network structure of multivariate time series." In: Scientific Reports 5, pp. 15508+.

Ladekarl, J., Peters, E., and Yie, J. (2019). Risk Cascades: Anticipating Resilient and Fragile Markets. Tech. rep. First Quadrant.

Laidlaw, J. (2019). Amid global push for sustainable investing, AI could be crucial catalyst. Tech. rep. S &P Global.
Lam, H. and Li, F. (2020). "Parametric Scenario Optimization under Limited Data: A Distributionally Robust Optimization View." In: arXiv e-Print.
Lamperti, F., Roventini, A., and Sani, A. (2018). "Agent-based model calibration using machine learning surrogates." In: Journal of Economic Dynamics and Control 90, pp. 366–389.
Lamponi, D. (2015). "A Data-Driven Categorization of Investable Assets." In: The Journal of Investing 24(4), pp. 73–80.
Lane, H., Howard, C., and Hapke, H. (2019). Natural Language Processing in Action: Understanding, analyzing, and generating text with Python. Manning. 544 pp.
Lanza, A., Bernardini, E., and Faiella, I. (2020). "Mind the gap! Machine Learning, ESG Metrics and Sustainable Investment." In: SSRN e-Print.
Lapan, M. (2020). Deep Reinforcement Learning Hands-On. 2nd ed. Packt. 826 pp.
Laskin, M., Lee, K., Stooke, A., Pinto, L., Abbeel, P., and Srinivas, A. (2020). "Reinforcement Learning with Augmented Data." In: arXiv e-Print.
Laurinaityte, N., Meinerding, C., Schlag, C., and Thimme, J. (2019). "Elephants and the Cross-Section of Expected Returns." In: SSRN e-Print.
Lawyer, G. (2015). "Understanding the influence of all nodes in a network." In: Scientific Reports 5(1), pp. 8665+.
Le Fur, E. and Outreville, J.-F. (2019). "Fine wine returns: a review of the literature." In: Journal of Asset Management 20, pp. 196–214.
Lee, M.-C., Lin, J.-C., and Gran, E. G. (2020). "ReRe: A Lightweight Real-time Ready-to-Go Anomaly Detection Approach for Time Series." In: arXiv e-Print.
Lee, W. (2012). "Risk On/Risk Off." In: The Journal of Portfolio Management 38(3), pp. 28–39.
Lee, W.-H., Ortiz, J., Ko, B., and Lee, R. (2018). "Time Series Segmentation through Automatic Feature Learning." In: arXiv e-Print.

Lemieux, V., Rahmdel, P. S., Walker, R., Wong, B. L. W., and Flood, M. (2014). "Clustering Techniques And Their Effect on Portfolio Formation and Risk Analysis." In: Proceedings of the International Workshop on Data Science for Macro-Modeling. DSMM'14. New York, NY, USA: ACM.

Leon, C., Kim, G.-Y., Martinez, C., and Lee, D. (2017a). "Equity markets' clustering and the global financial crisis." In: Quantitative Finance 17 (12), pp. 1905–1922.

Leon, D., Aragon, A., Sandoval, J., Hernandez, G., Arevalo, A., and Nino, J. (2017b). "Clustering algorithms for Risk-Adjusted Portfolio Construction." In: Procedia Computer Science 108, pp. 1334–1343.

Lettau, M. and Pelger, M. (2020). "Factors That Fit the Time Series and Cross-Section of Stock Returns." In: The Review of Financial Studies 33(5), pp. 2274–2325.

Leung, E., Lohre, H., Mischlich, D., Shea, Y., and Stroh, M. (2020). "The Promises and Pitfalls of Machine Learning for Predicting Cross-Sectional Stock Returns." In: SSRN e-Print.

Lewin, M. and Campani, C. H. (2020). "Portfolio Management under Multiple Regimes: Strategies that Outperform the Market." In: Brazilian Review of Finance 18(3), p. 52.

Lezmi, E., Malongo, H., Roncalli, T., and Sobotka, R. (2018). "Portfolio Allocation with Skewness Risk: A Practical Guide." In: SSRN e-Print.

Lezmi, E., Roche, J., Roncalli, T., and Xu, J. (2020). Improving the Robustness of Trading Strategy Backtesting with Boltzmann Machines and Generative Adversarial Networks. Tech. rep. Amundi.

Li, B. and Rossi, A. G. (2021). "Selecting Mutual Funds from the Stocks They Hold: A Machine Learning Approach." In: SSRN e-Print.

Li, B. and Piqueira, N. (2019). "State-dependent size and value premium: evidence from a regime-switching asset pricing model." In: Journal of Asset Management 20, pp. 229–249.

Li, J., Sun, Y., Su, J., Suzuki, T., and Huang, F. (2020a). "Understanding Generalization in Deep Learning via Tensor Methods." In: arXiv e-Print.

Li, Q. and Zhang, Q. (2020). "A Unified Model for Financial Event Classification, Detection and Summarization." In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization.

Li, X., Deng, S., Wang, S., Lv, Z., and Wu, L. (2018). "Review of small data learning methods." In: IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). IEEE, pp. 106–109.

Li, Y., Jiang, X.-F., Tian, Y., Li, S.-P., and Zheng, B. (2019). "Portfolio optimization based on network topology." In: Physica A: Statistical Mechanics and its Applications 515, pp. 671–681.

Li, Y., Turkington, D., and Yazdani, A. (2020b). "Beyond the Black Box: An Intuitive Approach to Investment Prediction with Machine Learning." In: The Journal of Financial Data Science 2(1), pp. 61–75.

Li, Z., Liu, J., and Zhang, X. (2020c). "Similarity Measure of Multivariate Time Series Based on Segmentation." In: Proceedings of the 2020 12th International Conference on Machine Learning and Computing. ACM.

Liew, J. and Budavari, T. (2017). "The "Sixth" factor: A Social Media Factor Derived Directly from Tweet Sentiments." In: The Journal of Portfolio Management 43(3), pp. 102–111.

Lim, B. and Zohren, S. (2020). "Time Series Forecasting With Deep Learning: A Survey." In: arXiv e-Print.

Linden, I. van der, Haned, H., and Kanoulas, E. (2019). "Global Aggregations of Local Explanations for Black Box models." In: arXiv e-Print.

Lipton, A. and Lopez de Prado, M. (2020a). "Three Quant Lessons from COVID-19." In: SSRN e-Print.

Lipton, A. and Lopez de Prado, M. (2020b). "What quants can learn from the Covid crisis." In: Risk.

Lipus, R. and Smith, D. (2019). "Taming Big Data." In: Big data and machine learning in quantitative investment. Wiley, pp. 13–32.

Lisi, F. and Menardi, G. (2015). "Double clustering for rating mutual funds." In: Electronic Journal of Applied Statistical Analysis 8(1), pp. 44–56.

Lisi, F. and Otranto, E. (2010). "Clustering mutual funds by return and risk levels." In: Mathematical and Statistical Methods for Actuarial Sciences and Finance, pp. 183–191.

Little, M. A. and Badawy, R. (2020). "Causal bootstrapping." In: arXiv e-Print.

Liu, X., Chen, J., Nair, V., and Sudjianto, A. (2018). "Model Interpretation: A Unified Derivative-based Framework for Nonparametric Regression and Supervised Machine Learning." In: arXiv e-Print.

Liu, X. (2015). "Historical simulation with component weight and ghosted scenarios." In: Journal of Risk 18(11), pp. 1–25.

Liu, Z., Huynh, T. L. D., and Dai, P.-F. (2020). "Fears for COVID-19: The crash risk of stock market." In: arXiv e-Print.

Lohre, H., Rother, C., and Schafer, K. A. (2020). "Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-asset Multi-factor Allocations." In: Machine Learning for Asset Management: New Developments and Financial Applications. Ed. by E. Jurczenko. Wiley, pp. 329–368.

Loning, M., Bagnall, A., Ganesh, S., Kazakov, V., Lines, J., and Kiraly, F. J. (2019). "sktime: A Unified Interface for Machine Learning with Time Series." In: arXiv e-Print.

Loning, M. and Kiraly, F. (2020). "Forecasting with sktime: Designing sktime's New Forecasting API and Applying It to Replicate and Extend the M4 Study." In: arXiv e-Print.

Lopez de Prado, M. (2016). "Building Diversified Portfolios that Outperform Out of Sample." In: The Journal of Portfolio Management 42(4), pp. 59–69.

Lopez de Prado, M. (2018). Advances in Financial Machine Learning. Wiley. 400 pp.

Lopez de Prado, M. (2019). "Beyond econometrics: A roadmap towards financial machine learning." In: SSRN e-Print.

Lopez de Prado, M. (2020a). Machine learning for asset managers. Cambridge University Press. 190 pp.

Lopez de Prado, M. (2020b). “Overfitting: causes and solutions.” In: SSRN e-Print.

Lopez de Prado, M. (2020c). “Tactical Investment Algorithms.” In: Alternative Investment Analyst Review 9(1).

Loughran, T. and McDonald, B. (2020). “Textual Analysis in Finance.” In: Annual Review of Financial Economics 12(1), pp. 357–375.

Lowe, S., Madras, D., Zemel, R., and Welling, M. (2020). “Amortized Causal Discovery: Learning to Infer Causal Graphs from Time-Series Data.” In: arXiv e-Print.

Lu, Y. Y., Fan, Y., Lv, J., and Noble, W. S. (2018). “DeepPINK: reproducible feature selection in deep neural networks.” In: arXiv e-Print.

Luo, Y. (2020). “Natural Language Process and Machine Learning in Global Stock Selection.” In: Machine Learning for Asset Management: New Developments and Financial Applications. Wiley, pp. 191–230.

Lussange, J., Belianin, A., Bourgeois-Gironde, S., and Gutkin, B. (2020). “Learning and Cognition in Financial Markets: A Paradigm Shift for Agent-Based Models.” In: Advances in Intelligent Systems and Computing. Springer International Publishing, pp. 241–255.

Lyócsa, S., Baumohl, E., Výrost, T., and Molnár, P. (2020). “Fear of the coronavirus and the stock markets.” In: Finance Research Letters 36, p. 101735.

Lyócsa, S. and Molnár, P. (2020). “Stock market oscillations during the corona crash: The role of fear and uncertainty.” In: Finance Research Letters 36, p. 101707.

Ma, C., Jacobsen, B., and Lee, W. (2019). “Pest Control: Eliminating Nuisance Allocations through Empirical Asset Class Identification.” In: The Journal of Financial Data Science 1 (3), pp. 29–40.

Ma, F., Li, Y., Liu, L., and Zhang, Y. (2018). “Are low-frequency data really uninformative? A forecasting combination perspective.” In: The North American Journal of Economics and Finance 44, pp. 92–108.

Ma, H. (2015). "Adaptive Investment Approach." In: Alternative Investment Analyst Review 3(4).
Ma, T. and Zhang, A. (2019). "Incorporating Biological Knowledge with Factor Graph Neural Network for Interpretable Deep Learning." In: arXiv e-Print.
Madan, D. B. and Schoutens, W. (2018). "Selfsimilarity in long horizon asset returns." In: SSRN e-Print.
Maharaj, E. A., D'Urso, P., and Caiado, J. (2019). Time Series Clustering and Classification. CRC Press. 244 pp.
Maio, P. F. (2019). "Comparing Asset Pricing Models with Traded and Macro Risk Factors." In: SSRN e-Print.
Makhlouf, K., Zhioua, S., and Palamidessi, C. (2020). "On the Applicability of ML Fairness Notions." In: arXiv e-Print.
Makridakis, S., Hyndman, R. J., and Petropoulos, F. (2020). "Forecasting in social settings: The state of the art." In: International Journal of Forecasting 36 (1), pp. 15–28.
Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018a). "Statistical and Machine Learning forecasting methods: Concerns and ways forward." In: PLoS ONE 13(3), e0194889.
Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018b). "The M4 Competition: Results, findings, conclusion and way forward." In: International Journal of Forecasting 34(4), pp. 802–808.
Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2019). "The M4 Competition: 100,000 time series and 61 forecasting methods." In: International Journal of Forecasting 36 (1), pp. 54–74.
Malik, M. M. (2020). "A Hierarchy of Limitations in Machine Learning." In: arXiv e-Print.
Man, X., Luo, T., and Lin, J. (2019). "Financial Sentiment Analysis (FSA): A Survey." In: IEEE International Conference on Industrial Cyber Physical Systems.
Mancuso, A. C. and Werner, L. (2013). "Review of combining forecasts approaches." In: Independent Journal of Management and Production 4(1).
Mannix, R. (2019). "Quants clone private equity." In: Risk.
Mannix, R. (2020). "Investors turn to raw data over ratings in ESG alpha hunt." In: Risk.

Marcinkevics, R. and Vogt, J. E. (2020). "Interpretability and Explainability: A Machine Learning Zoo Mini-tour." In: arXiv e-Print.

Mariani, G., Zhu, Y., Li, J., Scheidegger, F., Istrate, R., Bekas, C., and Malossi, A. C. I. (2019). "PAGAN: Portfolio Analysis with Generative Adversarial Networks." In: arXiv e-Print.

Marrs, T. (2019). "Introduction to Matrix Profiles: A Novel Data Structure for Mining Time Series." In: Towards Data Science.

Marti, G. (2020). "CORRGAN: sampling realistic financial correlation matrices using generative adversarial networks." In: ICASSP - IEEE nternational Conference on Acoustics, Speech and Signal Processing. IEEE, pp. 8459–8463.

Marti, G., Nielsen, F., Bihkowski, M., and Donnat, P. (2020). "A review of two decades of correlations, hierarchies, networks and clustering in financial markets." In: arXiv e-Print.

Martin-Utrera, A. (2020). "Shrinking Against Sentiment: Exploiting Behavioral Biases in Portfolio Optimization." In: SSRN e-Print.

Masih, R. (2012). Regime-Based Asset Allocation. Tech. rep. BNY Mellon Asset Management.

Massaro, P., Vannini, I., and Giudice, O. (2020). "Institutional Sector Classifier, a Machine Learning Approach." In: SSRN e-Print.

Mastakouri, A. A., Scholkopf, B., and Janzing, D. (2020). "Necessary and sufficient conditions for causal feature selection in time series with latent common causes." In: arXiv e-Print.

McDonald, S., Coleman, S., McGinnity, T. M., Li, Y., and Belatreche, A. (2014). "A comparison of forecasting approaches for capital markets." In: IEEE Conference on Computational Intelligence for Financial Engineering and Economics (CIFEr). London, UK: IEEE, pp. 32–39.

McMahan, B. and Rao, D. (2019). Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning. O'Reilly Media. 256 pp.

Mehta, D., Desai, D., and Pradeep, J. (2020). "Machine Learning Fund Categorizations." In: arXiv e-Print.

Messmer, M. and Audrino, F. (2017). "The (Adaptive) Lasso in the Zoo - Firm Characteristic Selection in the Cross-Section of Expected Returns." In: SSRN e-Print.

Messmer, M. and Audrino, F. (2020). "The Lasso and the Factor Zoo - Expected Returns in the Cross-Section." In: SSRN e-Print.

Meucci, A. (2009). "Enhancing the Black Litterman and related approaches: Views and stress-test on risk factors." In: Journal of Asset Management 10(2), pp. 89–96.

Meucci, A. (2012). A Fully Integrated Liquidity and Market Risk Model. Tech. rep. EDHEC.

Milutinovic, M., Schoenfeld, B., Ray, S., Shah, S., and Yan, D. (2020). "On Evaluation of AutoML Systems." In: 7th ICML Workshop on Automated Machine Learning.

Mireshghallah, F., Taram, M., Vepakomma, P., Singh, A., Raskar, R., and Esmaeilzadeh, H. (2020). "Privacy in Deep Learning: A Survey." In: arXiv e-Print.

Mishra, K., Basu, S., and Maulik, U. (2020). "A segmentation based similarity measure for time series data." In: Proceedings of the 7th ACM IKDD CoDS and 25th COMAD. ACM.

Mizuta, T. (2019). "An Agent-Based Model for Designing a Financial Market That Works Well." In: SSRN e-Print.

Mohammad, S. M. (2020). "Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text." In: arXiv e-Print.

Mohseni, S., Zarei, N., and Ragan, E. D. (2018). "A Survey of Evaluation Methods and Measures for Interpretable Machine Learning." In: arXiv e-Print.

Molnar, C. (2020a). Interpretable Machine Learning. 320 pp.

Molnar, C. (2020b). Interpretable Machine Learning: A Guide for Making Black Box Models Explainable. 312 pp.

Molnar, C., Konig, G., Herbinger, J., Freiesleben, T., Dandl, S., Scholbeck, C. A., Casalicchio, G., Grosse-Wentrup, M., and Bischl, B. (2020). "Pitfalls to Avoid when Interpreting Machine Learning Models." In: arXiv e-Print.

Molyboga, M. (2020). “A Modified Hierarchical Risk Parity Framework for Portfolio Management.” In: The Journal of Financial Data Science 2(3), pp. 128–139.

Monokroussos, G. and Zhao, Y. (2020). “Nowcasting in real time using popularity priors.” In: International Journal of Forecasting 36(3), pp. 1173–1180.

Montero-Manso, P., Athanasopoulos, G., Hyndman, R. J., and Talagala, T. S. (2020). “FFORMA: Feature-based Forecast Model Averaging.” In: International Jurnal of Forecasting 36 (1), pp. 86–92.

Moore, J., Hammerla, N., and Watkins, C. (2019). “Explaining Deep Learning Models with Constrained Adversarial Examples.” In: PRICAI 2019: trends in artificial intelligence. Ed. by A. C. Nayak and A. Sharma. Vol. 11670. Springer International Publishing, pp. 43–56.

Moraffah, R., Karami, M., Guo, R., Raglin, A., and Liu, H. (2020). “Causal Interpretability for Machine Learning - Problems, Methods and Evaluation.” In: ACM SIGKDD Explorations Newsletter.

Moreira, C., Chou, Y.-L., Velmurugan, M., Ouyang, C., Sindhgatta, R., and Bruza, P. (2020). “An Interpretable Probabilistic Approach for Demystifying Black-box Predictive Models.” In: arXiv e-Print.

Mullainathan, S. and Spiess, J. (2017). “Machine Learning: An Applied Econometric Approach.” In: Journal of Economic Perspectives 31(2), pp. 87–106.

Mulvey, J. M. and Liu, H. (2016). “Identifying Economic Regimes: Reducing Downside Risks for University Endowments and Foundations.” In: The Journal of Portfolio Management 43(1), pp. 100–108.

Munnix, M. C., Schafer, R., and Grothe, O. (2014). “Estimating correlation and covariance matrices by weighting of market similarity.” In: Quantitative Finance 14(5), pp. 931–939.

Munnix, M. C., Shimada, T., Schafer, R., Leyvraz, F., Seligman, T. H., Guhr, T., and Stanley, H. E. (2012). “Identifying States of a Financial Market.” In: Scientific Reports 2(1).

Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R., and Yu, B. (2019). "Interpretable machine learning: definitions, methods, and applications." In: arXiv e-Print.

Murtagh, F. and Contreras, P. (2012). "Algorithms for hierarchical clustering: an overview." In: WIREs: Data Mining and Knowledge Discovery 2(1), pp. 86–97.

Murtagh, F. and Contreras, P. (2017). "Algorithms for hierarchical clustering: an overview, II." In: WIREs: Data Mining and Knowledge Discovery 7(6), e1219–n/a.

Nachum, O., Tang, H., Lu, X., Gu, S., Lee, H., and Levine, S. (2019). "Why Does Hierarchy (Sometimes) Work So Well in Reinforcement Learning?" In: arXiv e-Print.

Nakagawa, K., Abe, M., and Komiyama, J. (2019a). "A Robust Transferable Deep Learning Framework for Cross-sectional Investment Strategy." In: arXiv e-Print.

Nakagawa, K., Ito, T., Abe, M., and Izumi, K. (2019b). "Deep Recurrent Factor Model: Interpretable Non-Linear and Time-Varying Multi-Factor Model." In: arXiv e-Print.

Nalewaik, J. (2015). Regime-Switching Models for Estimating Inflation Uncertainty. Tech. rep. Federal Reserve Board.

Narabin, S. and Boongasame, L. (2018). "A Cluster Analysis of Mutual Funds Data." In: International Conference on Big Data and Artificial Intelligence.

Nauman, B. (2019). "Artificial intelligence promises to enhance sustainable investing." In: Financial Times.

Naumer, H.-J. and Yurtoglu, B. B. (2020). "Good news, bad news, no news: the media and the cross section of stock returns." In: SSRN e-Print.

Ni, C., Li, Y., Forsyth, P., and Carroll, R. (2020a). "Optimal Asset Allocation for Outperforming a Stochastic Benchmark Target." In: SSRN e-Print.

Ni, H., Szpruch, L., Wiese, M., Liao, S., and Xiao, B. (2020b). "Conditional Sig-Wasserstein GANs for Time Series Generation." In: arXiv e-Print.

Nietert, B. and Otto, T. (2020). "Empirical asset pricing: economic significance and economic model evaluation." In: SSRN e-Print.

Nogueira, S., Sechidis, K., and Gavin, G. (2018). "On the Stability of Feature Selection Algorithms." In: Journal of Machine Learning Research 18(174), pp. 1–54.

Noguer i Alonso, M. and Srivastava, S. (2020). "Deep Reinforcement Learning for Asset Allocation in US Equities." In: arXiv e-Print.

Nori, H., Jenkins, S., Koch, P., and Caruana, R. (2019). "InterpretML: A Unified Framework for Machine Learning Interpretability." In: arXiv e-Print.

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.-E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., Kompatsiaris, I., Kinder-Kurlanda, K., Wagner, C., Karimi, F., Fernandez, M., Alani, H., Berendt, B., Kruegel, T., Heinze, C., Broelemann, K., Kasneci, G., Tiropanis, T., and Staab, S. (2020). "Bias in Data-driven AI Systems – An Introductory Survey." In: arXiv e-Print.

Nunes, M., Gerding, E., McGroarty, F., and Niranjan, M. (2020). "Long short-term memory networks and laglasso for bond yield forecasting: Peeping inside the black box." In: arXiv e-Print.

Nystrup, P., Hansen, B. W., Larsen, H. O., Madsen, H., and Lindstrom, E. (2018a). "Dynamic Allocation or Diversification: A Regime-Based Approach to Multiple Assets." In: The Journal of Portfolio Management 44(2), pp. 62–73.

Nystrup, P., Hansen, B. W., Madsen, H., and Lindstrom, E. (2015). "Regime-Based Versus Static Asset Allocation: Letting the Data Speak." In: The Journal of Portfolio Management 42(1), pp. 103–109.

Nystrup, P., Lindstrom, E., and Madsen, H. (2020). "Hyperparameter Optimization for Portfolio Selection." In: The Journal of Financial Data Science 22(3) (3), pp. 40–54.

Nystrup, P., Madsen, H., and Lindstrom, E. (2018b). "Dynamic portfolio optimization across hidden market regimes." In: Quantitative Finance 18(1), pp. 83–95.

Odendahl, F., Rossi, B., and Sekhposyan, T. (2020). "Comparing Forecast Performance with State Dependence." In: SSRN e-Print.

Oh, C., Han, S., and Jeong, J. (2020). "Time-Series Data Augmentation based on Interpolation." In: Procedia Computer Science 175, pp. 64–71.

Oneto, L. and Chiappa, S. (2020). "Fairness in Machine Learning." In: Recent Trends in Learning From Data. Springer International Publishing, pp. 155–196.

Ontanon, S. (2020). "An overview of distance and similarity functions for structured data." In: Artificial Intelligence Review 53(7), pp. 5309–5351.

Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio, Y. (2019). "N-BEATS: Neural basis expansion analysis for interpretable time series forecasting." In: arXiv e-Print.

Oreshkin, B. N., Carpov, D., Chapados, N., and Bengio, Y. (2020). "Meta-learning framework with applications to zero-shot time-series forecasting." In: arXiv e-Print.

Orozco, B. P., Abbati, G., and Roberts, S. (2018). "MOrdReD: Memory-based Ordinal Regression Deep Neural Networks for Time Series Forecasting." In: arXiv e-Print.

Overbeck, L. (2012). "Computational Issues in Stress Testing." In: Handbook of Computational Finance. Ed. by J.-C. Duan, W. K. Hardle, and J. E. Gentle. Springer Handbooks of Computational Statistics. Springer, pp. 651–673.

Ozbayoglu, A. M., Gudelek, M. U., and Sezer, O. B. (2020). "Deep Learning for Financial Applications : A Survey." In: Applied Soft Computing 93, p. 106384.

Paparrizos, J., Liu, C., Elmore, A. J., and Franklin, M. J. (2020). "Debunking Four Long-Standing Misconceptions of Time-Series Distance Measures." In: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data. ACM.

Papenbrock, J. and Schwendner, P. (2015). "Handling risk-on/risk-off dynamics with correlation regimes and correlation networks." In: Financial Markets and Portfolio Management 29(2), pp. 125–147.

Parafita, A. and Vitria, J. (2020). "Causal Inference with Deep Causal Graphs." In: arXiv e-Print.

Parker, F. J. (2016a). "Goal-Based Portfolio Optimization." In: The Journal of Wealth Management 19(3), pp. 22–30.

Parker, F. J. (2016b). "Portfolio Selection in a Goal-Based Setting." In: The Journal of Wealth Management 19(2), pp. 41–46.

Parker, F. J. (2020). "Allocation of wealth both within and across goals: a practitioner guide." In: The Journal of Wealth Management 23(1), pp. 8–21.

Patel, K. M. A. and Thakral, P. (2016). "The best clustering algorithms in data mining." In: International Conference on Communication and Signal Processing (ICCSP). IEEE, pp. 2042–2046.

Pearl, J. (2019). "The seven tools of causal inference, with reflections on machine learning." In: Communications of the ACM 62(3), pp. 54–60.

Pearl, J. and Mackenzie, D. (2018). The Book of Why. Basic Books. 432 pp.

Peralta, G. and Zareei, A. (2016). "A network approach to portfolio selection." In: Journal of Empirical Finance 38, pp. 157–180.

Pereiro, L. E. and Gonzalez-Rozada, M. (2015). "Forecasting Prices in Regime-Switching Markets." In: The Journal of Portfolio Management 41(4), pp. 133–139.

Perumal, K. and Flint, E. (2018). "Systematic testing of systematic trading strategies." In: Journal of Investment Strategies 7(3), pp. 29–49.

Pesenti, S. M., Bettini, A., Millossovich, P., and Tsanakas, A. (2020). "Scenario weights for importance measurement (SWIM) - an R package for sensitivity analysis." In: SSRN e-Print.

Pesenti, S. M., Millossovich, P., and Tsanakas, A. (2019). "Reverse sensitivity testing: What does it take to break the model?" In: European Journal of Operational Research 274(2), pp. 654–670.

Peters, E. (2015). "Stable vs. Unstable Markets: A Tale of Two States." In: SSRN e-Print.

Pharasi, H. K., Seligman, E., and Seligman, T. H. (2020). "Market states: A new understanding." In: arXiv e-Print.

Philipp, G. and Carbonell, J. G. (2018). "The Nonlinearity Coefficient - Predicting Overfitting in Deep Neural Networks." In: arXiv e-Print.

Pichler, A., Poledna, S., and Thurner, S. (2018). "Systemic-risk-efficient asset allocation: Minimization of systemic risk as a network optimization problem." In: arXiv e-Print.

Pike, T. and Vazquez-Grande, F. (2020). "Combining Forecasts: Can Machines Beat the Average?" In: SSRN e-Print.

Piliszek, R., Mnich, K., Migacz, S., Tabaszewski, P., Sulecki, A., Polewko-Klim, A., and Rudnicki, W. (2019). "MDFS - MultiDimensional Feature Selection." In: The R Journal.

Pinson, P. and Makridakis, S. (2021). "Pandemics and forecasting: The way forward through the Taleb-Ioannidis debate." In: International Journal of Forecasting, Early View.

Polson, N. and Sokolov, V. (2018a). "Deep Learning: Computational Aspects." In: arXiv e-Print.

Polson, N. G. and Sokolov, V. O. (2018b). "Deep Learning." In: arXiv e-Print.

Ponomareva, K., Roman, D., and Date, P. (2015). "An algorithm for moment-matching scenario generation with application to financial portfolio optimisation." In: European Journal of Operational Research 240(3), pp. 678–687.

Post, T., Karabati, S., and Arvanitis, S. (2019). "Robust optimization of forecast combinations." In: International Journal of Forecasting 35(3), pp. 910–926.

Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Vaughan, J. W., and Wallach, H. (2018). "Manipulating and Measuring Model Interpretability." In: arXiv e-Print.

Pozzi, F., Di Matteo, T., and Aste, T. (2013). "Spread of risk across financial markets: better to invest in the peripheries." In: Scientific Reports 3(1).

Preece, A. (2018). "Asking in AI: Explainability of intelligent systems - perspectives and challenges." In: Intelligent Systems in Accounting Finance & Management 25(2), pp. 63–72.

Probst, P., Boulesteix, A.-L., and Bischl, B. (2019). "Tunability: Importance of Hyperparameters of Machine Learning Algorithms." In: Journal of Mchine Learning Research 20(53), pp. 1–32.

Procacci, P. F., Phelan, C. E., and Aste, T. (2020). "Market structure dynamics during COVID-19 outbreak." In: arXiv e-Print.

Puerto, J., Rodriguez-Madrena, M., and Scozzari, A. (2020). "Clustering and portfolio selection problems: A unified framework." In: Computers & Operations Research 117, p. 104891.

Puiutta, E. and Veith, E. (2020). "Explainable Reinforcement Learning: A Survey." In: arXiv e-Print.

Pukthuanthong, K., Roll, R., and Subrahmanyam, A. (2019). "A Protocol for Factor Identification." In: The Review of Financial Studies 32(4), pp. 1573–1607.

Qi, G.-J. and Luo, J. (2019). "Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods." In: arXiv e-Print.

Qian, W., Rolling, C. A., Cheng, G., and Yang, Y. (2019). "On the forecast combination puzzle." In: Econometrics 7(3), p. 39.

Qian, X.-Y. (2017). "Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods." In: arXiv e-Print.

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., and Huang, X. (2020). "Pre-trained Models for Natural Language Processing: A Survey." In: arXiv e-Print.

Rafati, J. and Noelle, D. C. (2019). "Learning Representations in Model-Free Hierarchical Reinforcement Learning." In: Proceedings of the AAAI Conference on Artificial Intelligence 33, pp. 10009–10010.

Raffinot, T. (2017). "Hierarchical Clustering-Based Asset Allocation." In: The Journal of Portfolio Management 44(2), pp. 89–99.

Raffinot, T. (2018). "The Hierarchical Equal Risk Contribution Portfolio." In: SSRN e-Print.

Raghu, M. and Schmidt, E. (2020). "A Survey of Deep Learning for Scientific Discovery." In: arXiv e-Print.

Raman, N. and Leidner, J. L. (2019). "Financial Market Data Simulation Using Deep Intelligence Agents." In: Advances in Practical Applications of Survivable Agents and Multi-Agent Systems: The PAAMS Collection. Springer International Publishing, pp. 200–211.

Raman, N., Leidner, J. L., Vytelingum, K., and Horrell, G. (2019). Synthetic Reality: Synthetic market data generation at scale using agent based modeling. Tech. rep. Simudyne.

Rangapuram, S. S., Seeger, M. W., Gasthaus, J., Stella, L., Wang, Y., and Januschowski, T. (2018). "Deep State Space Models for Time Series Forecasting." In: Advances in Neural Information Processing Systems 31 (NIPS 2018).

Rapach, D. E. and Zhou, G. (2020). "Time-series and Cross-sectional Stock Return Forecasting: New Machine Learning Methods." In: Machine Learning for Asset Management: New Developments and Financial Applications. Ed. by E. Jurczenko. Wiley, pp. 1–33.

Raschka, S., Patterson, J., and Nolet, C. (2020). "Machine learning in Python: main developments and technology trends in data science, machine learning, and artificial intelligence." In: Information 11(4), p. 193.

Rasekhschaffe, K. and Jones, R. (2019). "Machine Learning for Stock Selection." In: Financial Analysts Journal 75 (3).

Ravula, S. (2021a). "Bankruptcy prediction using disclosure text features." In: arXiv e-Print.

Ravula, S. (2021b). "Text analysis in financial disclosures." In: arXiv e-Print.

Razafitombo, H. (2015). "Fund Analysis and Selection Based on the Dimensions of Performance Measures." In: The Journal of Index Investing 5(4), pp. 7–16.

Rebonato, R. (2017). "Stress Testing with Bayesian Nets and Related Techniques: Meeting the Engineering Challenges." In: The Palgrave handbook of unconventional risk transfer. Ed. by M. Pompella and N. A. Scordis. Springer International Publishing, pp. 537–575.

Rebonato, R. (2018). "The quickest way to lose the money you cannot afford to lose: reverse stress testing with maximum entropy." In: Journal of Risk 20(3), pp. 83–93.

Ren, P., Xiao, Y., Chang, X., Huang, P.-Y., Li, Z., Chen, X., and Wang, X. (2020). "A Survey of Deep Active Learning." In: arXiv e-Print.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016a). ""Why should I trust you?" Explaining the predictions of any classifier." In: Proceedings of 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, pp. 1135–1144.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016b). "Nothing Else Matters: Model-Agnostic Explanations By Identifying Prediction Invariance." In: arXiv e-Print.
Ribeiro, M. T., Singh, S., and Guestrin, C. G. (2018). "Anchors: High-Precision Model-Agnostic Explanations." In: 32nd AAAI Conference on Artificial Intelligence.
Riolo, M. A., Cantwell, G. T., Reinert, G., and Newman, M. E. J. (2017). "Efficient method for estimating the number of communities in a network." In: Physical Review E 96(3) (3), p. 032310.
Ritter, G. (2019). "Reinforcement learning in finance." In: Big data and machine learning in quantitative investment. Wiley, pp. 225–250.
Rocca, J. and Rocca, B. (2019). "Intuitively Understanding Variational Autoencoders." In: Towards Data Science.
Roder, F., Eppe, M., Nguyen, P. D. H., and Wermter, S. (2020). "Curious Hierarchical Actor-Critic Reinforcement Learning." In: arXiv e-Print.
Rodriguez-Barroso, N., Stipcich, G., Jimenez-Lopez, D., Ruiz-Millan, J. A., Martinez-Camara, E., Gonzalez-Seco, G., Luzon, M. V., Veganzones, M. A., and Herrera, F. (2020). "Federated Learning and Differential Privacy: Software tools analysis, the Sherpa.ai FL framework and methodological guidelines for preserving data privacy." In: Information Fusion 64(2), p. 270292.
Roncalli, T. (2019). "How Machine Learning Can Improve Portfolio Allocation of Robo-Advisors." In: SwissQuant Conference.
Roscher, R., Bohn, B., Duarte, M. F., and Garcke, J. (2019). "Explainable Machine Learning for Scientific Insights and Discoveries." In: arXiv e-Print.
Roscher, R., Bohn, B., Duarte, M. F., and Garcke, J. (2020). "Explainable Machine Learning for Scientific Insights and Discoveries." In: IEEE Access 8, pp. 42200–42216.
Rosen, D. (2015a). "Integrating Economic Scenarios with Advanced Scenario Analytics to Manage Investment Portfolios." In: SSRN e-Print.
Rosen, D. (2015b). "Re-Thinking Scenarios: Stress Testing of Multi-Asset Portfolios by Integrating Economic Scenarios with Advanced Simulation Analytics." In: SSRN e-Print.

Rosen, D. and Saunders, D. (2016). "Regress under stress: A simple least-squares method for integrating economic scenarios with risk simulations." In: Journal of Risk Management in Financial Institutions 9(4), pp. 391–412.

Rudin, C. (2019). "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead." In: Nature Machine Intelligence 1(5), pp. 206–215.

Ruenzi, S., Ungeheuer, M., and Weigert, F. (2020). "Joint Extreme events in equity returns and liquidity and their cross-sectional pricing implications." In: Journal of Banking & Finance 115, p. 105809.

Ruf, B., Boutharouite, C., and Detyniecki, M. (2020). "Getting Fairness Right: Towards a Toolbox for Practitioners." In: arXiv e-Print. arXiv: 2003.06920 [cs.AI].

Ruff, L., Kauffmann, J. R., Vandermeulen, R. A., Montavon, G., Samek, W., Kloft, M., Dietterich, T. G., and Muller, K.-R. (2020). "A Unifying Review of Deep and Shallow Anomaly Detection." In: arXiv e-Print.

Runge, J. (2018). "Causal network reconstruction from time series: From theoretical assumptions to practical estimation.." In: Chaos 28(7), p. 075310.

Ryll, L. and Seidens, S. (2019). "Evaluating the Performance of Machine Learning Algorithms in Financial Market Forecasting: A Comprehensive Survey." In: arXiv e-Print.

Saggioro, E., Wiljes, J. de, Kretschmer, M., and Runge, J. (2020). "Reconstructing regime-dependent causal relationships from observational time series." In: arXiv e-Print.

Sakakibara, T., Matsui, T., Mutoh, A., and Inuzuka, N. (2015). "Clustering Mutual Funds Based on Investment Similarity." In: Procedia Computer Science 60, pp. 881–890.

Saleiro, P., Rodolfa, K. T., and Ghani, R. (2020). "Dealing with Bias and Fairness in Data Science Systems." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Salhi, K., Deaconu, M., Lejay, A., Champagnat, N., and Navet, N. (2016). "Regime switching model for financial data: Empirical risk analysis." In: Physica A: Statistical Mechanics and its Applications 461, pp. 148–157.

Salinas, D., Flunkert, V., and Gasthaus, J. (2019). "DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks." In: arXiv e-Print.

Salinas, D., Flunkert, V., and Gasthaus, J. (2020). "DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks." In: International Journal of Forecasting 36 (3), pp. 1181–1191.

Salisu, A. A. and Akanni, L. O. (2020). "Constructing a Global Fear Index for the COVID-19 Pandemic." In: Emerging Markets Finance and Trade 56(10), pp. 2310–2331.

Samek, W., Montavon, G., Vedaldi, A., Hansen, L. K., and Muller, K.-R. (2019). Explainable AI: interpreting, explaining and visualizing deep learning. Ed. by W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Muller. Vol. 11700. Springer. 439 pp.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., and Muller, K.-R. (2020). "Toward Interpretable Machine Learning: Transparent Deep Neural Networks and Beyond." In: arXiv e-Print.

Samek, W. and Muller, K.-R. (2019). "Towards explainable artificial intelligence." In: Explainable AI: interpreting, explaining and visualizing deep learning. Ed. by W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Muller. Vol. 11700. Springer, pp. 5–22.

Samek, W., Wiegand, T., and Muller, K.-R. (2017). "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models." In: arXiv e-Print.

Sammut, C. and Webb, G. I. (2017). Encyclopedia of Machine Learning and Data Mining (Second Edition). Springer. 1352 pp.

Sanders, S. and Giraud-Carrier, C. (2017). "Informing the use of hyperparameter optimization through metalearning." In: IEEE International Conference on Data Mining (ICDM). IEEE, pp. 1051–1056.

Sandru, E.-D. and David, E. (2019). "Unified Feature Selection and Hyperparameter Bayesian Optimization for Machine Learning based Regression." In: International Symposium on Signals, Circuits and Systems. IEEE, pp. 1–5.

Sartorio, V. S. and Fonseca, T. C. O. (2020). "Dynamic clustering of time series data." In: arXiv e-Print.

Sato, Y. (2019). "Model-Free Reinforcement Learning for Financial Portfolios: A Brief Survey." In: arXiv e-Print.

Saxena, D. and Cao, J. (2020). "Generative Adversarial Networks (GANs): Challenges, Solutions, and Future Directions." In: arXiv e-Print.

Schmidt, P. and Biessmann, F. (2019). "Quantifying Interpretability and Trust in Machine Learning Systems." In: arXiv e-Print.

Schneider, J. and Handali, J. (2019). "Personalized explanation in machine learning: A conceptualization." In: arXiv e-Print.

Segal, S., Adi, Y., Pinkas, B., Baum, C., Ganesh, C., and Keshet, J. (2020). "Fairness in the Eyes of the Data: Certifying Machine-Learning Models." In: arXiv e-Print.

Seib, V., Lange, B., and Wirtz, S. (2020). "Mixing Real and Synthetic Data to Enhance Neural Network Training – A Review of Current Approaches." In: arXiv e-Print.

Sezer, O. B., Gudelek, M. U., and Ozbayoglu, A. M. (2020). "Financial time series forecasting with deep learning : A systematic literature review: 2005-2019." In: Applied Soft Computing 90, p. 106181.

Shalett, L. (2015). An Outcomes-Oriented Approach to Alternatives. Tech. rep. Morgan Stanley Wealth Management.

Sharma, S., Henderson, J., and Ghosh, J. (2019). "CERTIFAI: Counterfactual Explanations for Robustness, Transparency, Interpretability, and Fairness of Artificial Intelligence models." In: arXiv e-Print.

Shaub, D. (2020). "Fast and accurate yearly time series forecasting with forecast combinations." In: International Journal of Forecasting 33(1), pp. 116–120.
Sheikh, A. Z. and Sun, J. (2012). "Regime Change: Implications of Macroeconomic Shifts on Asset Class and Portfolio Performance." In: The Journal of Investing 21(3), pp. 36–54.
Shen, S., Zhu, T., Wu, D., Wang, W., and Zhou, W. (2021). "From distributed machine learning to federated learning: In the view of data privacy and security." In: Concurrency and Computation: Practice and Experience, Early View.
Shi, Q. and Li, B. (2021). "The evaluation and comparison of three benchmark asset pricing models with daily data: supplementary evidence." In: Asia-Pacific Journal of Accounting & Economics, Early View.
Shi, W., Song, S., Wang, Z., and Huang, G. (2020). "Self-Supervised Discovering of Causal Features: Towards Interpretable Reinforcement Learning." In: arXiv e-Print.
Shiono, T. (2020). "Estimation of Agent-Based Models Using Bayesian Deep Learning Approach of BayesFlow." In: SSRN e-Print.
Siami-Namini, S. and Namin, A. S. (2018). "Forecasting Economics and Financial Time Series: ARIMA vs. LSTM." In: arXiv e-Print.
Siddique, A., Hasan, I., and Lynch, D. (2019). Stress Testing (Second Edition). Risk Books. 500 pp.
Sidi, L. (2020). "Improving S&P stock prediction with time series stock similarity." In: arXiv e-Print.
Simonian, J. and Wu, C. (2019a). "Factors in Time: Fine-Tuning Hedge Fund Replication." In: The Journal of Portfolio Management 45 (3), pp. 159–164.
Simonian, J. and Wu, C. (2019b). "Minsky vs. Machine: New Foundations for Quant-Macro Investing." In: The Journal of Financial Data Science 1(2), pp. 94–110.
Singh, J. and Anand, A. (2019). "EXS: explainable search using local model agnostic interpretability." In: Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining - WSDM '19. ACM Press, pp. 770–773.

Sirignano, J. and Cont, R. (2018). "Universal features of price formation in financial markets: perspectives from deep learning." In: Quantitative Finance 19 (9), pp. 1449–1459.

Sironi, P. (2015). Modern Portfolio Management: from Markowitz to Probabilistic Scenario Optimisation. Risk Books. 204 pp.

Sirotyuk, E. (2019). "State of machine learning applications in investment management." In: Big data and machine learning in quantitative investment. Wiley, pp. 33–49.

Skoglund, J. (2019). "Quantification of model risk in stress testing and scenario analysis." In: Journal of Risk Model Validation 13(1), pp. 1–23.

Smyl, S. (2020). "A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting." In: International Journal of Forecasting 36(1) (1), pp. 75–85.

Snow, D. (2020). "Machine Learning in Asset Management - Part 2: Portfolio Construction - Weight Optimization." In: The Journal of Financial Data Science 2 (2), pp. 17–24.

Sokolov, A., Mostovoy, J., Ding, J., and Seco, L. (2021). "Building Machine Learning Systems for Automated ESG Scoring." In: The Journal of Impact and ESG Investing, Early View.

Sornette, D. (2014). "Dragon-kings and Predictions: Diagnostics and Forecasts for the World Financial Crisis." In: SSRN e-Print.

Sornette, D. and Cauwels, P. (2015). "Financial Bubbles: Mechanisms and Diagnostics." In: Review of Behavioral Economics 2(3), pp. 279–305.

Spiliotis, E., Abolghasemi, M., Hyndman, R. J., Petropoulos, F., and Assimakopoulos, V. (2020). "Hierarchical forecast reconciliation with machine learning." In: arXiv e-Print.

Spinner, T., Schlegel, U., Schafer, H., and El-Assady, M. (2020). "explAIner: A Visual Analytics Framework for Interactive and Explainable Machine Learning.." In: IEEE Transactions on Visualization and Computer Graphics 26 (1), pp. 1064–1074.

Sreevani and Murthy, C. A. (2017). "Bridging Feature Selection and Extraction: Compound Feature Generation." In: IEEE Transactions on Knowledge and Data Engineering 29(4), pp. 757–770.

SSgA Research (2015). Optimizing asset allocations to market regimes. Tech. rep. State Street Global Advisors.

Stanczyk, U. and Jain, L. C. (2015). Feature selection for data and pattern recognition. Ed. by U. Stanczyk and L. C. Jain. Studies in computational intelligence. Springer. 355 pp.

Stewart, M. (2020). "Guide to Interpretable Machine Learning." In: TopBots.

Suarez, J. L., Garcia, S., and Herrera, F. (2020). "pyDML: A Python Library for Distance Metric Learning." In: Journal of Machine Learning Research 21(96), pp. 1–7.

Suhonen, A., Lennkh, M., and Perez, F. (2017). "Quantifying Backtest Overfitting in Alternative Beta Strategies." In: The Journal of Portfolio Management 43 (2), pp. 90–104.

Sun, Q., Liu, Y., Chen, Z., Chua, T.-S., and Schiele, B. (2019). "Meta-Transfer Learning through Hard Tasks." In: arXiv e-Print.

Sutton, R. S. and Barto, A. G. (2018). Reinforcement Learning: An Introduction (Second Edition). MIT Press. 552 pp.

Takahashi, S., Chen, Y., and Tanaka-Ishii, K. (2019). "Modeling financial time-series with generative adversarial networks." In: Physica A: Statistical Mechanics and its Applications 527, p. 121261.

Taleb, N. N., Bar-Yam, Y., and Cirillo, P. (2021). "On single point forecasts for fat-tailed variables." In: International Journal of Forecasting, Early View.

Tang, X., Hu, F., and Wang, P. (2018). "Out-of-sample equity premium prediction: A scenario analysis approach." In: Journal of Forecasting 37(5), pp. 604–626.

Taylor, L. and Nitschke, G. (2017). "Improving Deep Learning using Generic Data Augmentation." In: arXiv e-Print.

Teiletche, J., Blin, O., Ielpo, F., and Lee, J. (2017). A macro risk-based approach to alternative risk premia allocation. Tech. rep. Unigestion.

Thompson, N. C., Greenewald, K., Lee, K., and Manso, G. F. (2020). "The Computational Limits of Deep Learning." In: arXiv e-Print.

Thomson, M. E., Pollock, A. C., Onkal, D., and Gonul, M. S. (2019). "Combining forecasts: Performance and coherence." In: International Journal of Forecasting 35(2), pp. 474–484.

Tola, V., Lillo, F., Gallegati, M., and Mantegna, R. N. (2008). "Cluster analysis for portfolio optimization." In: Journal of Economic Dynamics and Control 32(1), pp. 235–258.

Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavvaf, N., and Fox, E. A. (2020). "Natural Language Processing Advancements By Deep Learning: A Survey." In: arXiv e-Print.

Traccucci, P., Dumontier, L., Garchery, G., and Jacot, B. (2019). "A Triptych Approach for Reverse Stress Testing of Complex Portfolios." In: Risk (Cutting Edge).

Tran, M., Ngo, M., Pham-Hi, D., and Bui, M. (2020). "Bayesian Calibration of Hyperparameters in Agent-Based Stock Market." In: RIVF International Conference on Computing and Communication Technologies (RIVF). IEEE.

Tran, T., Pham, T., Carneiro, G., Palmer, L., and Reid, I. (2017). "A Bayesian Data Augmentation Approach for Learning Deep Models." In: arXiv e-Print.

Tran, V. L. (2021). "Sentiment and Covariance Characteristics." In: SSRN e-Print.

Triastcyn, A. and Faltings, B. (2020). "Bayesian Differential Privacy for Machine Learning." In: arXiv e-Print.

Tsiakas, I., Li, J., and Zhang, H. (2020). "Equity premium prediction and the state of the economy." In: Journal of Empirical Finance 58, pp. 75–95.

Tu, H. and Nair, V. (2018). "Is One Hyperparameter Optimizer Enough?" In: arXiv e-Print.

Tuggener, L., Amirian, M., Rombach, K., Lorwald, S., Varlet, A., Westermann, C., and Stadelmann, T. (2019). "Automated Machine Learning in Practice: State of the Art and Recent Results." In: 6th Swiss Conference on Data Science (SDS).

Udrescu, S.-M., Tan, A., Feng, J., Neto, O., Wu, T., and Tegmark, M. (2020). "AI Feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity." In: arXiv e-Print.

Udrescu, S.-M. and Tegmark, M. (2020). "AI Feynman: A physics-inspired method for symbolic regression." In: Science Advances 6(16), eaay2631.

Upreti, B. R., Back, P. M., Malo, P., Ahlgren, O., and Sinha, A. (2019). "Knowledge-Driven Approaches for Financial News Analytics." In: Network Theory and Agent-Based Modeling in Economics and Finance. Ed. by A. S. Chakrabarti, L. Pichl, and T. Kaizoji. Springer, pp. 375–404.

van Beek, M. (2020). "Consistent Calibration of Economic Scenario Generators: The Case for Conditional Simulation." In: arXiv e-Print.

van den Burg, G. J. J. and Williams, C. K. I. (2020). "An Evaluation of Change Point Detection Algorithms." In: arXiv e-Print.

van Rijn, J. N. and Hutter, F. (2018). "Hyperparameter importance across datasets." In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM Press, pp. 2367–2376.

Vasic, M., Petrovic, A., Wang, K., Nikolic, M., Singh, R., and Khurshid, S. (2019). "MoET: Interpretable and Verifiable Reinforcement Learning via Mixture of Expert Trees." In: arXiv e-Print.

Vaughan, J., Sudjianto, A., Brahimi, E., Chen, J., and Nair, V. N. (2018). "Explainable Neural Networks based on Additive Index Models." In: arXiv e-Print.

Vieira, A. (2020). "Generating Synthetic Sequential Data using GANs." In: Towards AI.

Vigen, T. (2019). Spurious Correlations. URL: https://www.tylervigen.com/spurious-correlations.

Vincent, K., Hsu, Y.-C., and Lin, H.-W. (2020). "Investment styles and the multiple testing of cross-sectional stock return predictability." In: Journal of Financial Markets, Early View.

Virgolin, M., Alderliesten, T., and Bosman, P. A. N. (2019). "On Explaining Machine Learning Models by Evolving Crucial and Compact Features." In: arXiv e-Print.

Vo, H. T. and Maurer, R. (2013). "Dynamic Asset Allocation under Regime Switching, Predictability and Parameter Uncertainty." In: SSRN e-Print.

Volpi, R., Namkoong, H., Sener, O., Duchi, J., Murino, V., and Savarese, S. (2018). "Generalizing to Unseen Domains via Adversarial Data Augmentation." In: arXiv e-Print.

Vozlyublennaia, N. and Wu, Y. (2018). "Mutual Funds Apart from the Crowd." In: SSRN e-Print.

Vrontos, S. D., Galakis, J., and Vrontos, I. D. (2021). "Modeling and predicting U.S. recessions using machine learning techniques." In: International Journal of Forecasting, Early View.

Vyrost, T., Lyocsa, S., and Baumohl, E. (2019). "Network-based asset allocation strategies." In: The North American Journal of Economics and Finance 47, pp. 516–536.

Walker, R. (2019). "Can AI resolve ESG rating shortfalls?" In: Expert Investor.

Waller, N. G. (2020). "Generating Correlation Matrices With Specified Eigenvalues Using the Method of Alternating Projections." In: The American Statistician 74(1), pp. 21–28.

Wang, H., Suri, A., Laster, D., and Almadi, H. (2011). "Portfolio Selection in Goals-Based Wealth Management." In: The Journal of Wealth Management 14(1), pp. 55–65.

Wang, R., Nie, K., Chang, Y.-J., Gong, X., Wang, T., Yang, Y., and Long, B. (2020a). "Deep Learning for Anomaly Detection." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Wang, Y., Polson, N. G., and Sokolov, V. O. (2019a). "Scalable Data Augmentation for Deep Learning." In: arXiv e-Print.

Wang, Y., Menkovski, V., Wang, H., Du, X., and Pechenizkiy, M. (2020b). "Causal Discovery from Incomplete Data: A Deep Learning Approach." In: arXiv e-Print.

Wang, Y., Smola, A., Maddix, D., Gasthaus, J., Foster, D., and Januschowski, T. (2019b). "Deep Factors for Forecasting." In: Proceedings of Machine Learning Research 97, pp. 6607–6617.

Wen, Q., Sun, L., Song, X., Gao, J., Wang, X., and Xu, H. (2020). "Time Series Data Augmentation for Deep Learning: A Survey." In: arXiv e-Print.

Wiese, M., Knobloch, R., Korn, R., and Kretschmer, P. (2020). "Quant GANs: deep generation of financial time series." In: Quantitative Finance, Early View.

Wigglesworth, R. (2019). "Quant funds train sights on private equity market." In: Financial Times.

Wolchover, N. (2018). "Machine Learning's 'Amazing' Ability to Predict Chaos." In: Quanta Magazine.

Wolff, D. and Echterling, F. (2020). "Stock Picking with Machine Learning." In: SSRN e-Print.
Wong, J. C. (2020). "Computational Causal Inference." In: arXiv e-Print.
Wu, R. and Keogh, E. J. (2020). "Current Time Series Anomaly Detection Benchmarks are Flawed and are Creating the Illusion of Progress." In: arXiv e-Print.
Wu, H.-S. (2016). "A survey of research on anomaly detection for time series." In: 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). Chengdu, China: IEEE, pp. 426–431.
Xie, N., Ras, G., Gerven, M. van, and Doran, D. (2020). "Explainable Deep Learning: A Field Guide for the Uninitiated." In: arXiv e-Print.
Xie, Q., Dai, Z., Hovy, E., Luong, M.-T., and Le, Q. V. (2019). "Unsupervised Data Augmentation." In: arXiv e-Print.
Xing, L. (2019). "Learning and Exploiting Multiple Subgoals for Fast Exploration in Hierarchical Reinforcement Learning." In: arXiv e-Print.
Xu, K.-L. (2020). "Testing for Multiple-Horizon Predictability: Direct Regression Based versus Implication Based." In: The Review of Financial Studies 33(9), pp. 4403–4443.
Yacoby, Y., Pan, W., and Doshi-Velez, F. (2020). "Characterizing and Avoiding Problematic Global Optima of Variational Autoencoders." In: Proceedings of Machine Learning Research, pp. 1–17.
Yan, G., Jia, S., Ding, J., Xu, X., and Pang, Y. (2019). "A time series forecasting based on cloud model similarity measurement." In: Soft computing 23, pp. 5443–5454.
Yan, X. and Bien, J. (2018). "Rare Feature Selection in High Dimensions." In: arXiv e-Print.
Yang, H., Lee, H. J., Cho, E., and Cho, S. (2016). "Automatic classification of securities using hierarchical clustering of the 10-Ks." In: IEEE International Conference on Big Data.
Yang, L. and Shami, A. (2020). "On Hyperparameter Optimization of Machine Learning Algorithms: Theory and Practice." In: arXiv e-Print.
Yao, H., Jia, X., Kumar, V., and Li, Z. (2020a). "Learning with Small Data." In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM.

Yao, L., Chu, Z., Li, S., Li, Y., Gao, J., and Zhang, A. (2020b). "A Survey on Causal Inference." In: arXiv e-Print.

Yaros, J. R. and Imielinski, T. (2015). "Data-driven methods for equity similarity prediction." In: Quantitative Finance 15(10), pp. 1657–1681.

Yazdani, A. (2020). "Machine Learning Prediction of Recessions: An Imbalanced Classification Approach." In: The Journal of Financial Data Science 2(4), pp. 21–32.

Yeh, C.-C. M. (2020). "Towards a Near Universal Time Series Data Mining Tool: Introducing the Matrix Profile." In: arXiv e-Print.

Yoon, J. (2020). "End-to-End Machine Learning Frameworks for Medicine: Data Imputation, Model Interpretation and Synthetic Data Generation." PhD thesis. UCLA.

Yoon, J., Jarrett, D., and van der Schaar, M. (2019). "Time-series Generative Adversarial Networks." In: NIPS. NIPS.

Yu, P. L. H., Li, W. K., and Ng, F. C. (2014). "Formulating hypothetical scenarios in correlation stress testing via a Bayesian framework." In: The North American Journal of Economics and Finance 27, pp. 17–33.

Yu, P. L. H., Ng, F. C., and Ting, J. K. W. (2020). "Adjusting covariance matrix for risk management." In: Quantitative Finance, Early View.

Yu, R. (2020). "A Tutorial on VAEs: From Bayes' Rule to Lossless Compression." In: arXiv e-Print.

Yu, T. and Zhu, H. (2020). "Hyper-Parameter Optimization: A Review of Algorithms and Applications." In: arXiv e-Print.

Zaffaroni, P. (2019). "Factor Models for Asset Pricing." In: SSRN e-Print.

Zakamulin, V. (2020). "Not All Bull and Bear Markets Are Alike: Insights From a Five-State Hidden Semi-Markov Model." In: SSRN e-Print.

Zanin, M., Papo, D., Sousa, P. A., Menasalvas, E., Nicchi, A., Kubik, E., and Boccaletti, S. (2016). "Combining complex networks and data mining: Why and how." In: Physics Reports 635, pp. 1–44.

Zaremba, A., Czapkiewicz, A., and Kambouris, G. D. (2020). "A Tale of Two States: An Application of a Markov Switching Model to Anomaly Returns." In: Eurasian Studies in Business and Economics. Springer International Publishing, pp. 227–240.

Zhang, A., Lipton, Z. C., Li, M., Smola, A. J., Werness, B., Hu, R., Zhang, S., and Tay, Y. (2020a). Dive into Deep Learning Book. 300 pp.

Zhang, C., Vinyals, O., Munos, R., and Bengio, S. (2018a). "A Study on Overfitting in Deep Reinforcement Learning." In: arXiv e-Print.

Zhang, H. and Yu, T. (2020). "Taxonomy of Reinforcement Learning Algorithms." In: Deep Reinforcement Learning. Springer Singapore, pp. 125–133.

Zhang, J. and Maringer, D. (2010). Asset Allocation under Hierarchical Clustering. Tech. rep. COMISEF Computational Optimization Methods in Statistics, Econometrics and Finance.

Zhang, J. and Maringer, D. (2011). "Distributing weights under hierarchical clustering: A way in reducing performance breakdown." In: Expert Systems with Applications 38(12), pp. 14952–14959.

Zhang, K. and Gencay, R. (2019). "Mutual fund performance in developing and advanced world networks." In: The Singapore Economic Review 64(02), pp. 399–421.

Zhang, Q. and Zhu, S.-C. (2018). "Visual Interpretability for Deep Learning: a Survey." In: arXiv e-Print.

Zhang, R., Li, W., and Mo, T. (2018b). "Review of Deep Learning." In: arXiv e-Print.

Zhang, X., Wang, Z., Liu, D., and Ling, Q. (2018c). "DADA: Deep Adversarial Data Augmentation for Extremely Low Data Regime Classification." In: arXiv e-Print.

Zhang, Y., Li, Z., and Zhang, Y. (2020b). "Validation and Calibration of an Agent-Based Model: A Surrogate Approach." In: Discrete Dynamics in Nature and Society 2020, pp. 1–9.

Zhang, Z., Tang, P., and Corpetti, T. (2020c). "Time Adaptive Optimal Transport: A Framework of Time Series Similarity Measure." In: IEEE Access 8, pp. 149764+.

Zhao, D., Zhang, L., Zhang, B., Zheng, L., Bao, Y., and Yan, W. (2019). "Deep Hierarchical Reinforcement Learning Based Recommendations via Multi-goals Abstraction." In: arXiv e-Print.

Zhao, L., Wang, G.-J., Wang, M., Bao, W., Li, W., and Stanley, H. E. (2018). "Stock market as temporal network." In: Physica A: Statistical Mechanics and its Applications 506, pp. 1104–1112.

Zhao, S. and Feng, Y. (2020). "For2For: Learning to forecast from forecasts." In: arXiv e-Print.

Zhou, X. (2019). Interpretability Methods in Machine Learning: A Brief Survey. Tech. rep. Two Sigma.

Zhou, Z. (2020). "Graphical Models in Financial Market and Portfolio Allocation: Applications and Considerations." PhD thesis. University of California Santa Barbara.

Zhu, Y., Gharghabi, S., Silva, D. F., Dau, H. A., Yeh, C.-C. M., Senobari, N. S., Almaslukh, A., Kamgar, K., Zimmerman, Z., Funning, G., Mueen, A., and Keogh, E. (2020). "The Swiss army knife of time series data mining: ten useful things you can do with the matrix profile and ten lines of code." In: Data Mining and Knowledge Discovery 34(4), pp. 949–979.

Zoller, M.-A. and Huber, M. F. (2020). "Benchmark and Survey of Automated Machine Learning Frameworks." In: arXiv e-Print.