References with abstracts for $\mathbb{QWIM}$ project: Reinforcement learning in quantitative wealth and investment management

Cristian Homescu

December 2022

Abstract

This document includes the list of references (including abstracts) for this QWIM project

Contents

1 Motivation for the project

1.1 Reinforcement learning

Reinforcement learning $\mathbb{RL}$ considers the problem of the automatic learning of optimal decisions over time. It differs from other types of learning (such as unsupervized or supervised) since the learning follows from feedback and experience (and not from some fixed training sample of data).

$\mathbb{RL}$ is learning what to do (how to map situations to actions), to maximize a numerical reward signal. The learner is not told which actions to take, but instead must discover which actions yield the most reward by trying them. Actions may affect not only the immediate reward but also the next situation and, through that, all subsequent rewards. These two characteristics (trial-and-error search and delayed reward) are the two most important distinguishing features of reinforcement learning.

Thus $\mathbb{RL}$ algorithms describe how an agent can learn an optimal action policy in a sequential decision process, through repeated experience, with primary purpose of finding an optimal policy (a mapping from the states of the world to the set of actions), in order to maximize cumulative reward (a long term strategy), although exploring might be sub-optimal on a short-term horizon.

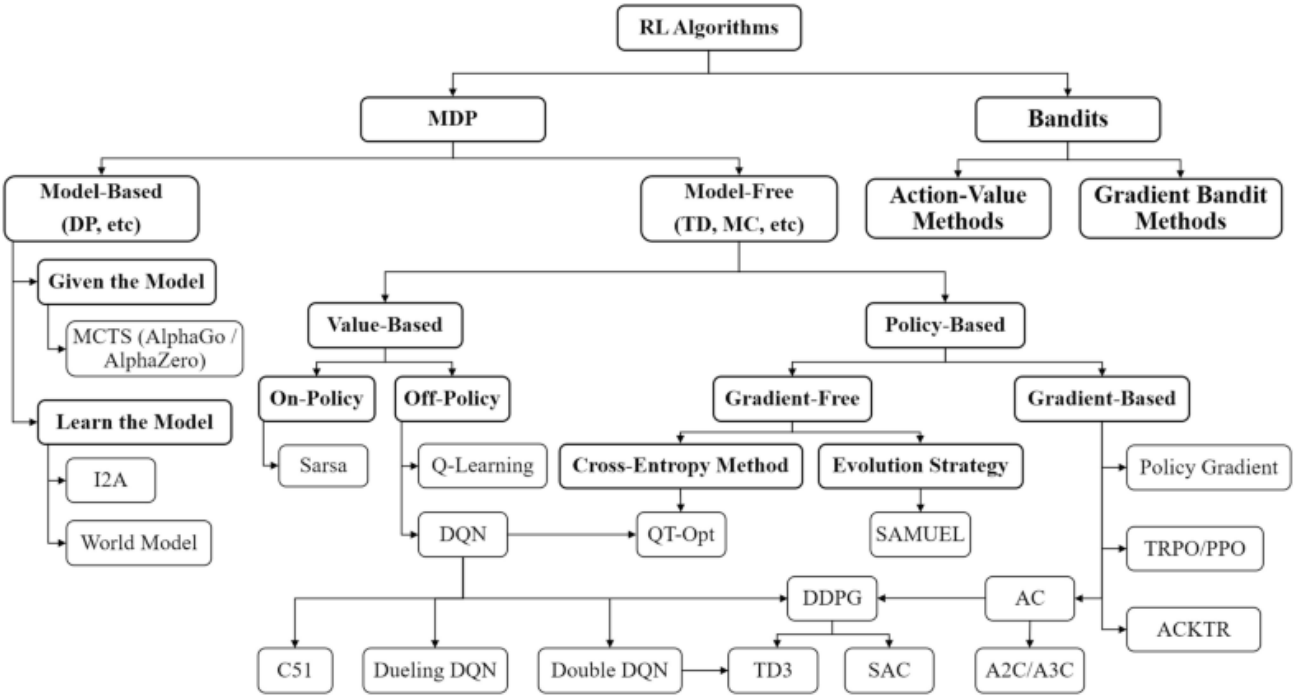
In a given environment, the agent policy provides some intermediate and terminal rewards, while the agent learns sequentially. When an agent picks an action, she can not infer ex-post the rewards induced by other action choices. Agent's actions have consequences, influencing not only rewards, but also future states of the world.

$\mathbb{RL}$ combines the following:

- major entities
  - ◇ agent: somebody or something who/that interacts with the environment by executing certain actions, making observations, and receiving eventual rewards for this. It is the learner and decision maker
  - ◇ environment: everything outside of an agent's direct control
  - ◇ policy: mapping from perceived states of the environment to actions to be taken when in those states, which defines how an agent selects actions
  - ◇ (optional) model of environment: mimics the behavior of the environment, or more generally, that allows inferences to be made about how the environment will behave
- communication channels:
  - ◇ actions: executed by the agent and given to the environment
  - ◇ reward: describes the goal of a $\mathbb{RL}$ problem, and it is a scalar value (obtained from environment) whose main purpose is to tell our agent how well it has behaved
  - ◇ value function: value of a state is the total amount of reward an agent can expect to accumulate over the future, starting from that state. Whereas the reward signal indicates what is good in an immediate sense, a value function specifies what is good in the long run
  - ◇ observations: information besides the reward that the agent receives from the environment, indicating what's going on around the agent

A taxonomy of reinforcement learning algorithms is shown next. Boxes with thick lines denote different categories, others denote specific algorithms.

Figure 1: Taxonomy of einforcement learning algorithms



Source: Zhang and Yu ("Taxonomy of Reinforcement Learning Algorithms," 2020)

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

$\mathbb{RL}$ allows to combine "prediction" and "portfolio construction" tasks in one integrated step, thereby closely aligning $\mathbb{ML}$ problem with investor objectives. $\mathbb{RL}$ is capable of handling real-world complexity of financial planning, including effects of income taxes, mean-reverting asset classes, time-varying bond yield curves, and uncertain life expectancies. $\mathbb{RL}$ allows robo-advisor to "learn" investor risk preference over time by observing her portfolio choices under different markets.

- Das and Varma ("Dynamic Goals-Based Wealth Management using Reinforcement Learning," 2020): use $\mathbb{RL}$ to solve for Goal Based Investing $\mathbb{GBI}$, with the solution converging to the solution obtained using dynamic programming
- Dixon and Halperin ("Goal-based wealth management with generative reinforcement learning," 2021), Dixon and Halperin ("G-Learner and GIRL: Goal Based Wealth Management with Reinforcement Learning," 2020), Dixon and Halperin ("G-Learner and GIRL: Goal Based Wealth Management with Reinforcement Learning," 2020): $\mathbb{GBI}$ approach using $\mathbb{RL}$:
  - ◇ G-Learner, which operates with explicitly defined one-step rewards, does not assume a data generation process and is suitable for noisy data
  - ◇ 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
- Fischer (*Reinforcement learning in financial markets - a survey*, 2018): $\mathbb{RL}$ combines "prediction" and "portfolio construction"
- Noguer i Alonso and Srivastava ("Deep Reinforcement Learning for Asset Allocation in US Equities," 2020): $\mathbb{DRL}$ for model-free asset allocation
- Sato ("Model-Free Reinforcement Learning for Financial Portfolios: A Brief Survey," 2019): model-free $\mathbb{RL}$ for financial portfolios

- Irlam ("Machine learning for retirement planning," 2020): $\mathbb{RL}$ is capable of handling real-world complexity of financial planning; taxes, inflation, longevity, etc.
- Irlam ("Machine learning for retirement planning," 2020): financial and retirement planning via $\mathbb{DRL}$ outperforms traditional approaches
- Benhamou et al. ("Adaptive learning for financial markets mixing model-based and model-free RL for volatility targeting," 2021): bridges the gap between $\mathbb{DRL}$ and Modern Portfolio Theory, with $\mathbb{DRL}$ mapping directly market conditions to actions by design
- Benhamou et al. ("Adaptive learning for financial markets mixing model-based and model-free RL for volatility targeting," 2021): augmented asset management via $\mathbb{DRL}$
- Hu and Lin ("Deep reinforcement learning for optimizing finance portfolio management," 2019), Cong et al. ("AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI," 2022): $\mathbb{DRL}$-based portfolio management
- Cong et al. ("Deep Sequence Modeling: Development and Applications in Asset Pricing," 2021), Cong et al. ("AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI," 2022): $\mathbb{DRL}$ plus deep sequence modeling delivers AlphaPortfolio with impressive out-of-sample performance.
- Alsabah et al. ("Robo-Advising: Learning Investors Risk Preferences via Portfolio Choices," 2021): using $\mathbb{RL}$, roboadvisor "learns" over time investor risk preference by observing her portfolio choices under different markets.
- Cao et al. ("Deep Hedging of Derivatives Using Reinforcement Learning," 2020): $\mathbb{DRL}$ for optimal hedging under transaction costs
- Carbonneau ("Deep Hedging of Long-Term Financial Derivatives," 2020), Carbonneau and Godin ("Equal risk pricing of derivatives with deep hedging," 2021), Cao et al. ("Deep Hedging of Derivatives Using Reinforcement Learning," 2021), Benhamou et al. ("Time your hedge with Deep Reinforcement Learning," 2020): $\mathbb{DRL}$ for hedging of financial derivatives (including long term)

Dixon and Halperin ("Goal-based wealth management with generative reinforcement learning," 2021): $\mathbb{RL}$ does not require a specific parametric model for the asset price dynamics, offering instead a data-driven extension of dynamic programming. However, practical applications of $\mathbb{RL}$ to problems of portfolio optimization amount to a high-dimensional control in a continuous space of allocations (actions, in the language of $\mathbb{RL}$). However, methods based on deep $\mathbb{RL}$ are very data-hungry, slow to train, very sensitive to both the inputs and hyperparameters, and usually rely heavily on various heuristics.

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

Goal Based Investing $\mathbb{GBI}$ is an investment approach where performance is measured by the success of investments in meeting the investor financial goals. For an individual investor such goals may be retirement, education for children, or vacation home, while for insitutional investors such as pension funds the goals may be aligned to the pension liabilities. The objective is to invest systematically in a consistent manner with investor's risk profile and time horizon of the goals, and performance is measured by probabilities of success in achieving investor financial goals (at given time horizons and for specified priority levels)

Portfolio optimization for GBI can be viewed as an optimal control problem performed within a data-driven world. $\mathbb{GBI}$ can be solved via reinforcement learning $\mathbb{RL}$, a paradigm of learning by trial-and-error, solely from rewards or punishments. It was shown that $\mathbb{RL}$ can solve financial applications of intertemporal choice.

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. One possible solution may leverage hierarchical $\mathbb{RL}$ and curiosity-driven exploration.

1.4 Using $\mathbb{RL}$ in $\mathbb{QWIM}$: Asset allocation and portfolio construction

Benhamou ("Next Generation Robo Advisors," 2021)

Benhamou et al. ("AAMDRL: Augmented Asset Management With Deep Reinforcement Learning," 2021)

Benhamou et al. ("Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning," 2021)

Benhamou et al. ("Deep Reinforcement Learning (DRL) for Portfolio Allocation," 2021)

Benhamou et al. ("DRLPS: Deep Reinforcement Learning for Portfolio Selection (Presentation Slides ECML PKDD)," 2021)

Benhamou et al. ("Deep Reinforcement Learning (DRL) for Portfolio Allocation," 2021)

Benhamou et al. ("Bridging the gap between Markowitz planning and deep reinforcement learning (ICAPS PRL Presentation Slides 2020)," 2021)

Ungari and Benhamou ("Deep Reinforcement Learning for Portfolio Allocation," 2021)

2 Relevant references

2.1 Main references

List of references:

Aliaga-Diaz et al. (*Vanguard's Life-Cycle Investing Model (VLCM): A general portfolio framework for goals-based investing*, 2021)

Alsabah et al. ("Robo-Advising: Learning Investors Risk Preferences via Portfolio Choices," 2021)

Bartram et al. ("Machine Learning for Active Portfolio Management," 2021)

Benhamou et al. ("AAMDRL: Augmented Asset Management With Deep Reinforcement Learning," 2021)

Benhamou et al. ("Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning," 2021)

Benhamou et al. ("Deep Reinforcement Learning (DRL) for Portfolio Allocation," 2021)

Benhamou et al. ("Detecting and Adapting to Crisis Pattern with Context Based Deep Reinforcement Learning," 2021)

Benhamou et al. ("Adaptive learning for financial markets mixing model-based and model-free RL for volatility targeting," 2021)

Benhamou et al. ("Bridging the gap between Markowitz planning and deep reinforcement learning (ICAPS PRL Presentation Slides 2020)," 2021)

Cong et al. ("AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI," 2022)

Cuschieri et al. ("TD3-Based Ensemble Reinforcement Learning for Financial Portfolio Optimisation," 2021)

Das et al. ("Dynamic portfolio allocation in goals-based wealth management," 2020)

Das and Varma ("Dynamic Goals-Based Wealth Management using Reinforcement Learning," 2020)

Das et al. ("Optimal Goals-Based Investment Strategies For Switching Between Bull and Bear Markets," 2021)

Das et al. ("Dynamic portfolio allocation in goals-based wealth management," 2020)

Das et al. ("Optimal Goals-Based Investment Strategies For Switching Between Bull and Bear Markets," 2022)

Das et al. ("Dynamic optimization for multi-goals wealth management," 2022)

Dempster et al. ("Lifecycle Goal Achievement or Portfolio Volatility Reduction?" 2016)

Dixon and Halperin ("G-Learner and GIRL: Goal Based Wealth Management with Reinforcement Learning," 2020)

Dixon and Halperin ("Goal-based wealth management with generative reinforcement learning," 2021)

Dixon et al. (*Machine Learning in Finance: from theory to practice*, 2020)

Dong et al. ("Baconian: A Unified Open source Framework for Model-Based Reinforcement Learning," 2021)

Fleiss et al. ("Deep Reinforcement Learning and Feature Extraction For Constructing Alpha Generating Equity Portfolios," 2021)

Guan and Liu ("Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach," 2021)

Halperin et al. ("Combining Reinforcement Learning and Inverse Reinforcement Learning for Asset Allocation Recommendations," 2022)

Hambly et al. ("Recent Advances in Reinforcement Learning in Finance," 2021)

Honchar (*AI for portfolio management: from Markowitz to Reinforcement Learning*, 2019)

Irlam ("Machine learning for retirement planning," 2020)

Irlam ("Multi Scenario Financial Planning via Deep Reinforcement Learning AI," 2020)

Ivanov and D'yakonov ("Modern Deep Reinforcement Learning Algorithms," 2019)

Jaimungal et al. ("Robust Risk-Aware Reinforcement Learning," 2021)

Kim et al. ("Personalized goal-based investing via multi-stage stochastic goal programming," 2020)

Kolm and Ritter ("Modern Perspectives on Reinforcement Learning in Finance," 2020)

Liu et al. ("FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance," 2021)

Jaisson ("Deep differentiable reinforcement learning and optimal trading," 2022)

Martellini et al. ("Securing Replacement Income with Goal-Based Retirement Investing Strategies," 2020)

Marzban et al. ("WaveCorr: Correlation-savvy Deep Reinforcement Learning for Portfolio Management," 2021)

Mohammed et al. ("Embracing advanced AI/ML to help investors achieve success: Vanguard Reinforcement Learning for Financial Goal Planning," 2021)
Mulvey et al. ("A Factor- and Goal-Driven Model for Defined Benefit Pensions: Setting Realistic Benefits," 2019)
Noguer i Alonso and Srivastava ("Deep Reinforcement Learning for Asset Allocation in US Equities," 2020)
Nguyen et al. ("Review, Analyze, and Design a Comprehensive Deep Reinforcement Learning Framework," 2021)
Odermatt et al. ("Deep Reinforcement Learning for Finance and the Efficient Market Hypothesis," 2021)
Parker ("Goal-Based Portfolio Optimization," 2016)
Parker ("Portfolio Selection in a Goal-Based Setting," 2016)
Parker ("Allocation of wealth both within and across goals: a practitioner guide," 2020)
Parker ("Multi-Period Goals-Based Portfolio Optimization," 2021)
Parker ("Achieving Goals While Making an Impact: Balancing Financial Goals with Impact Investing," 2021)
Parker ("A Goals-Based Theory of Utility," 2021)
Plaat et al. ("High-Accuracy Model-Based Reinforcement Learning, a Survey," 2021)
Schwarzer et al. ("Pretraining Representations for Data-Efficient Reinforcement Learning," 2021)
Sun et al. ("Reinforcement Learning for Quantitative Trading," 2021)
Ungari and Benhamou ("Deep Reinforcement Learning for Portfolio Allocation," 2021)
Wang et al. ("Portfolio Selection in Goals-Based Wealth Management," 2011)
Yekelchik et al. ("Deep Q-Network Interpertability: Applications to ETF Trading," 2021)
Zhang et al. ("On the Importance of Hyperparameter Optimization for Model-based Reinforcement Learning," 2021)

2.2 Comprehensive list of references

2.2.1 Goals-based investing

List of references:

Brunel ("Extending the Goals-Based Framework to Comprise Both Investment and Financial Planning," 2020)
Chhabra (*The Aspirational Investor: Taming the Markets to Achieve Your Life's Goals*, 2015)
Cvitanic et al. ("Pi portfolio management: reaching goals while avoiding drawdowns," 2020)
Das et al. ("Dynamic portfolio allocation in goals-based wealth management," 2020)
Das and Varma ("Dynamic Goals-Based Wealth Management using Reinforcement Learning," 2020)
Das and Ross ("The Role of Options in Goals-Based Wealth Management," 2021)
Das et al. ("Dynamic portfolio allocation in goals-based wealth management," 2020)
Das et al. ("Optimal Goals-Based Investment Strategies For Switching Between Bull and Bear Markets," 2022)
Das et al. ("Dynamic optimization for multi-goals wealth management," 2022)
Deguest et al. (*Introducing a Comprehensive Investment Framework for Goals-Based Wealth Management*, 2015)
Dempster et al. ("Lifecycle Goal Achievement or Portfolio Volatility Reduction?" 2016)
Denault and Simonato ("A note on a dynamic goal-based wealth management problem," 2022)
Dixon et al. (*Machine Learning in Finance: from theory to practice*, 2020)
Dixon and Halperin ("G-Learner and GIRL: Goal Based Wealth Management with Reinforcement Learning," 2020)
Guo et al. ("Portfolio optimization with a prescribed terminal wealth distribution," 2022)
Janssen et al. ("Life Cycle Investing: From Target-Date to Goal-Based Investing," 2013)
Kim et al. ("Personalized goal-based investing via multi-stage stochastic goal programming," 2020)
Liu et al. ("FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance," 2021)
Martellini et al. ("Securing Replacement Income with Goal-Based Retirement Investing Strategies," 2020)
Mohammed et al. ("Embracing advanced AI/ML to help investors achieve success: Vanguard Reinforcement Learning for Financial Goal Planning," 2021)

Mulvey et al. ("A Factor- and Goal-Driven Model for Defined Benefit Pensions: Setting Realistic Benefits," 2019)

Muralidhar ("Asset Pricing, Asset Allocation and Risk-Adjusted Performance with Multiple Goals and Agency: The Goals and Risk-based Asset Pricing Model," 2020)

Parker ("Goal-Based Portfolio Optimization," 2016)

Parker ("Portfolio Selection in a Goal-Based Setting," 2016)

Parker ("Achieving Goals While Making an Impact: Balancing Financial Goals with Impact Investing," 2021)

Parker ("Multi-Period Goals-Based Portfolio Optimization," 2021)

Parker ("Allocation of wealth both within and across goals: a practitioner guide," 2020)

Parker ("A Goals-Based Theory of Utility," 2021)

Roncalli ("How Machine Learning Can Improve Portfolio Allocation of Robo-Advisors," 2019)

Roncalli ("Advanced Course in Asset Management," 2021)

Shalett (*An Outcomes-Oriented Approach to Alternatives*, 2015)

Wang et al. ("Portfolio Selection in Goals-Based Wealth Management," 2011)

2.2.2 Reinforcement learning within context of QWIM

List of references:

Aboussalah ("Topological Phase Space Reconstruction For Augmented Real-World Reinforcement Learning," 2021)

Aboussalah et al. ("What is the value of the cross-sectional approach to deep reinforcement learning?" 2022)

Alsabah et al. ("Robo-Advising: Learning Investors Risk Preferences via Portfolio Choices," 2021)

Andre and Coqueret ("Dirichlet policies for reinforced factor portfolios," 2021)

Ardon et al. ("Towards a fully RL-based Market Simulator," 2021)

Bartram et al. ("Machine Learning for Active Portfolio Management," 2021)

Benhamou et al. ("AAMDRL: Augmented Asset Management With Deep Reinforcement Learning," 2021)

Benhamou et al. ("Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning," 2021)

Benhamou et al. ("Deep Reinforcement Learning (DRL) for Portfolio Allocation," 2021)

Benhamou et al. ("Detecting and Adapting to Crisis Pattern with Context Based Deep Reinforcement Learning," 2021)

Benhamou et al. ("Adaptive learning for financial markets mixing model-based and model-free RL for volatility targeting," 2021)

Benhamou et al. ("Bridging the gap between Markowitz planning and deep reinforcement learning (ICAPS PRL Presentation Slides 2020)," 2021)

Borrageiro et al. ("Reinforcement Learning for Systematic FX Trading," 2021)

Borrageiro et al. ("The Recurrent Reinforcement Learning Crypto Agent," 2022)

Brini and Tantari ("Deep Reinforcement Trading with Predictable Returns," 2022)

Buehler et al. ("Deep hedging," 2019)

Carbonneau ("Deep Hedging of Long-Term Financial Derivatives," 2020)

Carbonneau and Godin ("Equal risk pricing of derivatives with deep hedging," 2021)

Cartea et al. ("Deep Reinforcement Learning for Algorithmic Trading," 2022)

Charpentier et al. ("Reinforcement Learning in Economics and Finance," 2020)

Coache and Jaimungal ("Reinforcement Learning with Dynamic Convex Risk Measures," 2022)

Cong et al. ("AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI," 2022)

Das et al. ("Dynamic portfolio allocation in goals-based wealth management," 2020)

Das and Varma ("Dynamic Goals-Based Wealth Management using Reinforcement Learning," 2020)

Das et al. ("Optimal Goals-Based Investment Strategies For Switching Between Bull and Bear Markets," 2021)

Das et al. ("Dynamic optimization for multi-goals wealth management," 2022)

Dixon and Halperin ("G-Learner and GIRL: Goal Based Wealth Management with Reinforcement Learning," 2020)

Dixon and Halperin ("Goal-based wealth management with generative reinforcement learning," 2021)

Dixon et al. (*Machine Learning in Finance: from theory to practice*, 2020)

Dong et al. ("Factor Representation and Decision Making in Stock Markets Using Deep Reinforcement Learning," 2021)

Du et al. ("Deep Reinforcement Learning for Option Replication and Hedging," 2020)

Fleiss et al. ("Deep Reinforcement Learning and Feature Extraction For Constructing Alpha Generating Equity Portfolios," 2021)

Forsyth et al. ("Optimal control of the decumulation of a retirement portfolio with variable spending and dynamic asset allocation," 2021)

Gasperov et al. ("Adaptive rolling window selection for minimum variance portfolio estimation based on reinforcement learning," 2020)

Gašperov et al. ("Reinforcement Learning Approaches to Optimal Market Making," 2021)

Gu ("Deep Reinforcement Learning with Function Properties in Mean Reversion Strategies," 2021)

Guan and Liu ("Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach," 2021)

Hambly et al. ("Recent Advances in Reinforcement Learning in Finance," 2021)

Hieu ("Deep Reinforcement Learning for Stock Portfolio Optimization," 2020)

Hirsa et al. ("Deep reinforcement learning on a multi-asset environment for trading," 2021)

Honchar (*AI for portfolio management: from Markowitz to Reinforcement Learning*, 2019)

Huang and Tanaka ("A Modularized and Scalable Multi-Agent Reinforcement Learning-based System for Financial Portfolio Management," 2021)

Huang et al. ("Deep reinforcement learning for portfolio management," 2022)

Irlam ("Machine learning for retirement planning," 2020)

Irlam ("Multi Scenario Financial Planning via Deep Reinforcement Learning AI," 2020)

Ivanov and D'yakonov ("Modern Deep Reinforcement Learning Algorithms," 2019)

Jaimungal et al. ("Robust Risk-Aware Reinforcement Learning," 2021)

Jaimungal ("Reinforcement learning and stochastic optimisation," 2022)

Jaisson ("Deep differentiable reinforcement learning and optimal trading," 2022)

Janner et al. ("Reinforcement Learning as One Big Sequence Modeling Problem," 2021)

Katongo and Bhattacharyya ("The Use of Deep Reinforcement Learning in Tactical Asset Allocation," 2021)

Liu et al. ("MTV: Visual Analytics for Detecting, Investigating, and Annotating Anomalies in Multivariate Time Series," 2021)

Li et al. ("FinRL-Podracer: High Performance and Scalable Deep Reinforcement Learning for Quantitative Finance," 2021)

Li ("An Automated Portfolio Trading System with Feature Preprocessing and Recurrent Reinforcement Learning," 2021)

Li ("Financial Trading with Feature Preprocessing and Recurrent Reinforcement Learning," 2021)

Li et al. ("FinRL-Podracer: High Performance and Scalable Deep Reinforcement Learning for Quantitative Finance," 2021)

Liao et al. ("Portfolio Allocation with Dynamic Risk Preference Via Reinforcement Learning: Evidence from the Taiwan 50 Index," 2022)

Liu et al. ("FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance," 2021)

Liu et al. ("Goal-Conditioned Reinforcement Learning: Problems and Solutions," 2022)

Ma et al. ("Deep Q-Learning for Trading Cryptocurrency," 2021)

Marzban et al. ("WaveCorr: Correlation-savvy Deep Reinforcement Learning for Portfolio Management," 2021)

Mohammed et al. ("Embracing advanced AI/ML to help investors achieve success: Vanguard Reinforcement Learning for Financial Goal Planning," 2021)

Mosavi et al. ("Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics," 2020)

Noguer i Alonso and Srivastava ("Deep Reinforcement Learning for Asset Allocation in US Equities," 2020)

Odermatt et al. ("Deep Reinforcement Learning for Finance and the Efficient Market Hypothesis," 2021)

Pricope ("Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review," 2021)

Soleymani and Paquet ("Deep Graph Convolutional Reinforcement Learning for Financial Portfolio Management - DeepPocket," 2021)

Sun et al. ("Reinforcement Learning for Quantitative Trading," 2021)

Suri et al. ("TradeR: Practical Deep Hierarchical Reinforcement Learning for Trade Execution," 2021)

Ungari and Benhamou ("Deep Reinforcement Learning for Portfolio Allocation," 2021)

Wang and Yu ("Robo-Advising: Enhancing Investment with Inverse Optimization and Deep Reinforcement Learning," 2021)

Wells and Bednarz ("Explainable AI and Reinforcement Learning—A Systematic Review of Current Approaches and Trends," 2021)

Yang et al. ("Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy," 2020)

Yashaswi ("Deep Reinforcement Learning for Portfolio Optimization using Latent Feature State Space (LFSS) Module," 2021)

2.2.3 Reinforcement learning

List of references:

Amin et al. ("A Survey of Exploration Methods in Reinforcement Learning," 2021)

Bareinboim (*Causal Reinforcement Learning*, 2020)

Chan et al. ("ESG in Factors," 2020)

Cruz Barsce et al. ("Automatic tuning of hyper-parameters of reinforcement learning algorithms using Bayesian optimization with behavioral cloning," 2021)

Dong et al. ("Baconian: A Unified Open source Framework for Model-Based Reinforcement Learning," 2021),

Dulac-Arnold et al. ("An empirical investigation of the challenges of real-world reinforcement learning," 2021)

Fedus et al. ("Hyperbolic Discounting and Learning over Multiple Horizons," 2019)

Francois-Lavet et al. ("An introduction to deep reinforcement learning," 2018)

Garau-Luis et al. ("Evaluating the progress of Deep Reinforcement Learning in the real world: aligning domain-agnostic and domain-specific research," 2021)

Glanois et al. ("A Survey on Interpretable Reinforcement Learning," 2022)

Hamadanian et al. ("Reinforcement Learning in Time-Varying Systems: an Empirical Study," 2022)

Hayes et al. ("A Practical Guide to Multi-Objective Reinforcement Learning and Planning," 2021)

Hoang et al. ("Successor Feature Landmarks for Long-Horizon Goal-Conditioned Reinforcement Learning," 2021)

Ivanov and D'yakonov ("Modern Deep Reinforcement Learning Algorithms," 2019)

Jordan et al. ("Evaluating the Performance of Reinforcement Learning Algorithms," 2020)

Khetarpal et al. ("Towards Continual Reinforcement Learning: A Review and Perspectives," 2021)

Lambert et al. ("Learning Accurate Long-term Dynamics for Model-based Reinforcement Learning," 2021)

Lapan (*Deep Reinforcement Learning Hands-On*, 2020)

Laskin et al. ("Reinforcement Learning with Augmented Data," 2020)

Laskin et al. ("URLB: Unsupervised Reinforcement Learning Benchmark," 2021)

Lazaridis et al. ("Deep Reinforcement Learning: A State-of-the-Art Walkthrough," 2020)

Lehnert and Littman ("Successor Features Combine Elements of Model-Free and Model-based Reinforcement Learning," 2020)

Li et al. ("Anchor: The achieved goal to replace the subgoal for hierarchical reinforcement learning," 2021)

Li et al. ("GenURL: A General Framework for Unsupervised Representation Learning," 2021)

Llorente et al. ("A survey of Monte Carlo methods for noisy and costly densities with application to reinforcement learning," 2021)

Moerland et al. ("Model-based Reinforcement Learning: A Survey," 2021)

Nachum et al. ("Why Does Hierarchy (Sometimes) Work So Well in Reinforcement Learning?" 2019)

Naeem et al. ("A Gentle Introduction to Reinforcement Learning and its Application in Different Fields," 2020)

Nguyen et al. ("Review, Analyze, and Design a Comprehensive Deep Reinforcement Learning Framework," 2021)

Parker-Holder et al. ("Automated Reinforcement Learning (AutoRL): A Survey and Open Problems," 2022)

Plaat ("Deep Reinforcement Learning," 2022)

Plaat et al. ("High-Accuracy Model-Based Reinforcement Learning, a Survey," 2021)
Qian and Yu ("Derivative-Free Reinforcement Learning: A Review," 2021)
Ren et al. ("A Free Lunch from the Noise: Provable and Practical Exploration for Representation Learning," 2021)
Schwarzer et al. ("Pretraining Representations for Data-Efficient Reinforcement Learning," 2021)
Shi et al. ("Self-Supervised Discovering of Causal Features: Towards Interpretable Reinforcement Learning," 2020)
Stapelberg and Malan ("A survey of benchmarking frameworks for reinforcement learning," 2021)
Sutton and Barto (*Reinforcement Learning: An Introduction (Second Edition)*, 2018)
Tarbouriech et al. ("Adaptive Multi-Goal Exploration," 2021)
Wang et al. ("Benchmarking Model-Based Reinforcement Learning," 2019)
Xing ("Learning and Exploiting Multiple Subgoals for Fast Exploration in Hierarchical Reinforcement Learning," 2019)
Yaghmaie and Ljung ("A Crash Course on Reinforcement Learning," 2021)
Zhang and Yu ("Taxonomy of Reinforcement Learning Algorithms," 2020)
Zhang et al. ("On the Importance of Hyperparameter Optimization for Model-based Reinforcement Learning," 2021)
Zhao et al. ("Deep Hierarchical Reinforcement Learning Based Recommendations via Multi-goals Abstraction," 2019)

2.2.4 Deep reinforcement learning

List of references:
Arulkumaran et al. ("Deep reinforcement learning: A brief survey," 2017)
Bertsekas ("Feature-Based Aggregation and Deep Reinforcement Learning: A Survey and Some New Implementations," 2018)
Cai et al. ("A Survey on Deep Reinforcement Learning for Data Processing and Analytics," 2022)
Chakraborty ("Deep Reinforcement Learning in Financial Markets," 2019)
Francois-Lavet et al. ("An introduction to deep reinforcement learning," 2018)
Fujita et al. ("ChainerRL: A Deep Reinforcement Learning Library," 2021)
Heuillet et al. ("Explainability in deep reinforcement learning," 2021)
Ivanov and D'yakonov ("Modern Deep Reinforcement Learning Algorithms," 2019)
Kirk et al. ("A Survey of Generalisation in Deep Reinforcement Learning," 2022)
Liu et al. ("Goal-Conditioned Reinforcement Learning: Problems and Solutions," 2022)
Majid et al. ("Deep Reinforcement Learning Versus Evolution Strategies: A Comparative Survey," 2021)
Pardo ("Tonic: A Deep Reinforcement Learning Library for Fast Prototyping and Benchmarking," 2021)
Rafati and Marcia ("Deep Reinforcement Learning via L-BFGS Optimization," 2018)
Raileanu et al. ("Automatic Data Augmentation for Generalization in Deep Reinforcement Learning," 2021)
Stooke and Abbeel ("Accelerated Methods for Deep Reinforcement Learning," 2018)
Stooke and Abbeel ("rlpyt: A Research Code Base for Deep Reinforcement Learning in PyTorch," 2019)
Sun et al. ("TimeTraveler: Reinforcement Learning for Temporal Knowledge Graph Forecasting," 2021)
Weng et al. ("Tianshou: a Highly Modularized Deep Reinforcement Learning Library," 2022)
Xiong et al. ("Practical Deep Reinforcement Learning Approach for Stock Trading," 2018)
Yang et al. ("Exploration in Deep Reinforcement Learning: A Comprehensive Survey," 2022)
Zhang et al. ("A Study on Overfitting in Deep Reinforcement Learning," 2018)
Zhu et al. ("Transfer Learning in Deep Reinforcement Learning: A Survey," 2021)

2.2.5 Inverse reinforcement learning

List of references:
Abbeel and Ng ("Inverse Reinforcement Learning," 2017)
Arora and Doshi ("A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress," 2018)
Brown et al. ("Risk-Aware Active Inverse Reinforcement Learning," 2019)
Brown and Niekum ("Machine teaching for inverse reinforcement learning: algorithms and applications," 2019)

Castro et al. ("Inverse Reinforcement Learning with Multiple Ranked Experts," 2019)

Freymuth et al. ("Versatile Inverse Reinforcement Learning via Cumulative Rewards," 2021)

Fu et al. ("Evaluating Strategic Structures in Multi-Agent Inverse Reinforcement Learning," 2021)

Halperin ("The QLBS Q-Learner goes NuQLear: fitted Q iteration, inverse RL, and option portfolios," 2019)

Halperin and Feldshteyn ("Market Self-Learning of Signals, Impact and Optimal Trading: Invisible Hand Inference with Free Energy (Or, How We Learned to Stop Worrying and Love Bounded Rationality)," 2018)

Krishnan et al. ("HIRL: Hierarchical Inverse Reinforcement Learning for Long-Horizon Tasks with Delayed Rewards," 2016)

2.2.6 Testing and comparison procedures for investment portfolios

References:

Adcock et al. ("Portfolio Performance Measurement: Monotonicity with Respect to the Sharpe Ratio and Multivariate Tests of Correlation," 2017)

Arnott et al. ("A backtesting protocol in the era of machine learning," 2019)

Bailey et al. ("Stock Portfolio Design and Backtest Overfitting," 2017)

Bessler and Wolff ("Portfolio Optimization with Industry Return Prediction Models," 2017)

Bessler et al. ("Multi-asset portfolio optimization and out-of-sample performance: an evaluation of Black Litterman, mean-variance, and naive diversification approaches," 2017)

Bjerring et al. ("Feature selection for portfolio optimization," 2017)

Bruni et al. ("On exact and approximate stochastic dominance strategies for portfolio selection," 2017)

Bruni et al. ("Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models," 2016)

Bryzgalova et al. ("Bayesian solutions for the factor zoo: we just ran two quadrillion models," 2021)

Cesarone et al. ("On the stability of portfolio selection models," 2019)

Cesarone et al. ("Why Small Portfolios Are Preferable and How to Choose Them," 2018)

Chaudhuri and Lo ("Dynamic Alpha: A Spectral Decomposition of Investment Performance Across Time Horizons," 2019)

Diris et al. ("Long-Term Strategic Asset Allocation: An Out-of-Sample Evaluation," 2015)

Fabozzi and Lopez de Prado ("Being Honest in Backtest Reporting: A Template for Disclosing Multiple Tests," 2018)

Greiner and Stoyanov ("Portfolio scoring by expected risk premium," 2019)

Guidolin et al. ("Portfolio performance of linear SDF models: an out-of-sample assessment," 2018)

Guo ("A Statistical Response to Challenges in Vast Portfolio Selection," 2019)

Guo et al. ("When Does The 1/N Rule Work?" 2019)

Haley ("K-fold cross validation performance comparisons of six naive portfolio selection rules: how naive can you be and still have successful out-of-sample portfolio performance?" 2017)

Harvey et al. ("An Evaluation of Alternative Multiple Testing Methods for Finance Applications," 2020)

Hens et al. ("Escaping the backtesting illusion," 2020)

Hsu et al. (*Do Cross-Sectional Stock Return Predictors Pass the Test without Data-Snooping Bias?* 2017)

Hsu et al. ("Asset allocation strategies, data snooping, and the 1 / N rule," 2018)

Huang and Yu ("A new procedure for resampled portfolio with shrinkaged covariance matrix," 2020)

Hwang et al. ("Naive versus optimal diversification: Tail risk and performance," 2018)

Ielpo et al. (*Engineering Investment Process: Making Value Creation Repeatable*, 2017)

Jaeger et al. ("Understanding machine learning for diversified portfolio construction by explainable AI," 2020)

Kazak and Pohlmeier ("Testing out-of-sample portfolio performance," 2019)

Kazak and Pohlmeier (*Portfolio Pretesting with Machine Learning*, 2020)

Kuntz ("Portfolio Strategies with Classical and Alternative Benchmarks," 2018)

Lohre et al. ("Hierarchical Risk Parity: Accounting for Tail Dependencies in Multi-asset Multi-factor Allocations," 2020)

Lopez de Prado ("A Data Science Solution to the Multiple-Testing Crisis in Financial Research," 2019)

Lopez de Prado and Lewis ("Detection of false investment strategies using unsupervised learning methods," 2019)

Malavasi et al. ("Second order of stochastic dominance efficiency vs mean variance efficiency," 2021)

Mooney et al. ("Dynamic Regime Strategy for Stress Testing and Optimizing Institutional Investor Portfolios," 2020)

Platanakis et al. ("Horses for Courses: Mean-Variance for Asset Allocation and 1/N for Stock Selection," 2021)

Radovanov and Marcikic ("Testing The Performance Of The Investment Portfolio Using Block Bootstrap Method," 2014)

Rebonato ("A financially justifiable and practically implementable approach to coherent stress testing," 2019)

Schumann ("Backtesting," 2019)

Seymour et al. ("Dynamic portfolio management strategies: A framework for historical analysis," 2018)

Suhonen et al. ("Quantifying Backtest Overfitting in Alternative Beta Strategies," 2017)

Taljaard and Maré ("Why has the equal weight portfolio underperformed and what can we do about it?" 2021)

Tayali ("A novel backtesting methodology for clustering in mean–variance portfolio optimization," 2020)

Traccucci et al. ("A Triptych Approach for Reverse Stress Testing of Complex Portfolios," 2019)

Valentine et al. ("Beyond p values: utilizing multiple methods to evaluate evidence," 2019)

Vincent et al. ("Analyzing the Performance of Multifactor Investment Strategies under a Multiple Testing Framework," 2018)

Vovk and Wang ("True and false discoveries with e-values," 2020)

Vovk and Wang ("E-values: Calibration, combination, and applications," 2021)

Wiecki et al. ("All That Glitters Is Not Gold: Comparing Backtest and Out-of-Sample Performance on a Large Cohort of Trading Algorithms," 2016)

Yu ("Comparing Classical Portfolio Optimization and Robust Portfolio Optimization on Black Swan Events," 2021)

Yuan and Zhou ("Why Naive 1/N Diversification Is Not So Naive, and How to Beat It?" 2022)

Zhang et al. ("DoubleEnsemble: A New Ensemble Method Based on Sample Reweighting and Feature Selection for Financial Data Analysis," 2020)

Zhang et al. ("Information Coefficient as a Performance Measure of Stock Selection Models," 2020)

Zhang et al. ("Deep Learning for Portfolio Optimization," 2020)

2.2.7 Software implementations and frameworks

List of references:

Bargiacchi et al. ("AI-Toolbox: A C++ library for Reinforcement Learning and Planning (with Python Bindings)," 2020)

Beysolow II (*Applied Reinforcement Learning with Python: With OpenAI Gym, Tensorflow, and Keras*, 2019)

Brockman et al. ("OpenAI Gym," 2016)

Bou and De Fabritiis ("PyTorchRL: Modular and Distributed Reinforcement Learning in PyTorch," 2021)

Castro et al. ("Dopamine: A Research Framework for Deep Reinforcement Learning," 2018)

Dong et al. ("Baconian: A Unified Open source Framework for Model-Based Reinforcement Learning," 2021)

D'Eramo et al. ("MushroomRL: Simplifying Reinforcement Learning Research," 2021)

Fujita et al. ("ChainerRL: A Deep Reinforcement Learning Library," 2021)

Hoffman et al. ("Acme: A Research Framework for Distributed Reinforcement Learning," 2020)

Hulbert et al. ("EasyRL: A Simple and Extensible Reinforcement Learning Framework," 2020)

Kolesnikov and Hrinchuk ("Catalyst.RL: A Distributed Framework for Reproducible RL Research," 2019)

Kuttler et al. ("TorchBeast: A PyTorch Platform for Distributed RL," 2019)

Li et al. ("RL: Generic reinforcement learning codebase in TensorFlow," 2019)

Liu et al. ("FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance," 2021)

Liu et al. ("Goal-Conditioned Reinforcement Learning: Problems and Solutions," 2022)

Loon et al. ("SLM Lab: A Comprehensive Benchmark and Modular Software Framework for Reproducible Deep Reinforcement Learning," 2019)

Nguyen et al. ("Review, Analyze, and Design a Comprehensive Deep Reinforcement Learning Framework," 2021)

Pardo ("Tonic: A Deep Reinforcement Learning Library for Fast Prototyping and Benchmarking," 2021)
Pineda et al. ("MBRL-Lib: A Modular Library for Model-based Reinforcement Learning," 2021)
Pretorius et al. ("Mava: a research framework for distributed multi-agent reinforcement learning," 2021)
Prollochs and Feuerriegel ("ReinforcementLearning: A Package to Perform Model-Free Reinforcement Learning in R," 2019)
Staley et al. ("The AI Arena: A Framework for Distributed Multi-Agent Reinforcement Learning," 2021)
Terry et al. ("PettingZoo: Gym for Multi-Agent Reinforcement Learning," 2021)
Weng et al. ("Tianshou: a Highly Modularized Deep Reinforcement Learning Library," 2022)
Xu and Chen ("A Validation Tool for Designing Reinforcement Learning Environments," 2021)

References

Abbeel, P. and Ng, A. Y. (2017). "Inverse Reinforcement Learning." In: *Encyclopedia of machine learning and data mining*. Ed. by C. Sammut and G. I. Webb. Boston, MA: Springer US, pp. 678–682.
Inverse reinforcement learning (inverse RL) considers the problem of extracting a reward function from observed (nearly) optimal behavior of an expert acting in an environment.

Aboussalah, A. M. (2021). "Topological Phase Space Reconstruction For Augmented Real-World Reinforcement Learning." In: *SSRN e-Print*.
The purpose of this paper is to develop a methodology to enhance solutions to real-world reinforcement learning (RWRL) problems based on time series data. We focus on portfolio management (PM) as a challenging real-world domain to test our approach. We demonstrate the utility of image-based data augmentation through empirical studies showing substantial improvements with higher cumulative returns and better convergence. We developed a theoretical framework using complexity bounds such as Rademacher complexities that explains the benefits of augmentation for RL. Our theoretical development contributes to a better fundamental understanding of why data augmentation works in the general RL setting even when we don't have an algebraic group structure (e.g. translations, reflections, rotations, etc.) as commonly used in most data augmentation techniques.

Aboussalah, A. M., Xu, Z., and Lee, C.-G. (2022). "What is the value of the cross-sectional approach to deep reinforcement learning?" In: *Quantitative Finance*, Early View.
Reinforcement learning (RL) for dynamic asset allocation is an emerging field of study. Total return, the common performance metric, is useful for comparing algorithms but does not help us determine how close an RL algorithm is to an optimal solution. In real-world financial applications, a bad decision could prove to be fatal. One of the key ideas of our work is to combine the two paradigms of the mean-variance optimization approach (Markowitz criteria) and the optimal capital growth approach (Kelly criteria) via the actor-critic approach. By using an actor-critic approach, we can balance optimization of risk and growth by configuring the actor to optimize the mean-variance while the critic is configured to maximize growth. We propose a Geometric Policy Score used by the critic to assess the quality of the actions taken by the actor. This could allow portfolio manager practitioners to better understand the investment RL policy. We present an extensive and in-depth study of RL algorithms for use in portfolio management (PM). We studied eight published policy-based RL algorithms which are preferred over value-based RL because they are better suited for continuous action spaces and are considered to be state of the art, Deterministic Policy Gradient (DPG), Stochastic Policy Gradients (SPG), Deep Deterministic Policy Gradient (DDPG), Trust Region Policy Optimization (TRPO), Proximal Policy Optimization (PPO), Twin Delayed Deep Deterministic Policy Gradient (TD3), Soft Actor Critic (SAC), and Evolution Strategies (ES) for Policy Optimization. We implemented all eight and we were able to modify all of them for PM but our initial testing determined that their performance was not satisfactory. Most algorithms showed difficulty converging during the training process due to the non-stationary and noisy nature of financial environments, along with other challenges. We selected the four most promising algorithms DPG, SPG, DDPG, PPO for further improvements. The modification of RL algorithms to finance required unconventional changes. We have developed a novel approach for encoding multi-type financial data in a way that is compatible with RL. We use a multi-channel convolutional neural network (CNN-RL) framework, where each channel corresponds to a specific type of data such as high-low-open-close prices and volumes. We also designed a reward function based on concepts such as alpha, beta, and diversification that are financially meaningful while still being learnable by RL. In addition, portfolio managers will typically use a blend of time series analysis and cross-sectional analysis before making a decision. We extend our approach to incorporate, for the first time, cross-sectional deep RL in addition to time series RL. Finally, we demonstrate the performance of the RL agents and benchmark them against commonly used passive and active trading strategies, such as the uniform buy-and-hold (UBAH) index and the dynamical multi-period Mean-Variance-Optimization (MVO) model.

Adcock, C., Areal, N., Armada, M. R., Cortez, M. C., Oliveira, B., and Silva, F. (2017). "Portfolio Performance Measurement: Monotonicity with Respect to the Sharpe Ratio and Multivariate Tests of Correlation." In: *SSRN e-Print*.
This paper reports an investigation into methods of portfolio performance measurement. The work is motivated first by equivocal empirical evidence reported by several authors about the correlation of performance measures with the Sharpe ratio. Secondly it is motivated by recent work which specifies that performance

measures will be monotone functions of the Sharpe ratio if portfolio returns follow the same location-scale distribution. The paper demonstrates that the class of location-scale distributions is broader than previously reported. It presents conditions under which monotonicity with respect to the Sharpe ratio will fail. The paper shows that for large sample sizes the correlation between pairs of performance measures that are functions of the Sharpe ratio is unity. The correct null hypothesis for tests of correlation is therefore rho=1. Two multivariate tests of this null hypothesis are presented. The new tests are used to carry out of a comprehensive study of performance measurement for a set over ninety UK investment trusts.

Aliaga-Diaz, R., Ahluwalia, H., Zhu, V., Donaldson, S., Daga, A., and Pakula, D. (2021). *Vanguard's Life-Cycle Investing Model (VLCM): A general portfolio framework for goals-based investing*. Tech. rep. Vanguard.
Investors have multiple goals throughout their lifetime, each requiring them to make complex, interconnected decisions about saving, spending, and asset allocation. We present a framework for making asset allocation decisions based on an investor's goals, preferences, and personal circumstances and factoring in the uncertainty of asset returns. The Vanguard Life-Cycle Investing Model (VLCM) is a proprietary model for glide-path construction that can assist in the creation of custom investment portfolios for retirement as well as nonretirement goals, such as saving for college. The VLCM embodies key principles of life-cycle investing theory, including a utility-based framework encompassing risk aversion and time preference. It also incorporates important behavioral finance considerations such as loss aversion and income shortfall aversion. The use of the VLCM enables cost-benefit analysis of glide-path customization, evaluation of risk-return trade-offs of various asset and sub-asset allocation choices, and multiple portfolio analytics of the probability of success and odds of income sufficiency. Based on VLCM's analytical framework, we find that risk-aversion levels are the dominant factor behind the broad stock-bond split in the glide path, affecting both glide-path slope and ending allocation.

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* 19(2), pp. 369–292.
We introduce a reinforcement learning framework for retail robo-advising. The robo-advisor does not know the investor risk preference but learns it over time by observing her portfolio choices in different market environments. We develop an exploration-exploitation algorithm that trades off costly solicitations of portfolio choices by the investor with autonomous trading decisions based on stale estimates of investor risk aversion. We show that the approximate value function constructed by the algorithm converges to the value function of an omniscient robo-advisor over a number of periods that is polynomial in the state and action space. By correcting for the investor mistakes, the robo-advisor may outperform a stand-alone investor, regardless of the investor opportunity cost for making portfolio decisions.

Amin, S., Gomrokchi, M., Satija, H., van Hoof, H., and Precup, D. (2021). "A Survey of Exploration Methods in Reinforcement Learning." In: *arXiv e-Print*.
Exploration is an essential component of reinforcement learning algorithms, where agents need to learn how to predict and control unknown and often stochastic environments. Reinforcement learning agents depend crucially on exploration to obtain informative data for the learning process as the lack of enough information could hinder effective learning. In this article, we provide a survey of modern exploration methods in (Sequential) reinforcement learning, as well as a taxonomy of exploration methods.

Andre, E. and Coqueret, G. (2021). "Dirichlet policies for reinforced factor portfolios." In: *arXiv e-Print*.
This article aims to combine factor investing and reinforcement learning (RL). The agent learns through sequential random allocations which rely on firms' characteristics. Using Dirichlet distributions as the driving policy, we derive closed forms for the policy gradients and analytical properties of the performance measure. This enables the implementation of REINFORCE methods, which we perform on a large dataset of US equities. Across a large range of parametric choices, our result indicates that RL-based portfolios are very close to the equally-weighted (1/N) allocation. This implies that the agent learns to be *agnostic* with regard to factors, which can partly be explained by cross-sectional regressions showing a strong time variation in the relationship between returns and firm characteristics.

Ardon, L., Vadori, N., Spooner, T., Xu, M., Vann, J., and Ganesh, S. (2021). "Towards a fully RL-based Market Simulator." In: *arXiv e-Print*.
We present a new financial framework where two families of RL-based agents representing the Liquidity Providers and Liquidity Takers learn simultaneously to satisfy their objective. Thanks to a parametrized reward formulation and the use of Deep RL, each group learns a shared policy able to generalize and interpolate over a wide range of behaviors. This is a step towards a fully RL-based market simulator replicating

complex market conditions particularly suited to study the dynamics of the financial market under various scenarios.

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.
Machine learning offers a set of powerful tools that holds considerable promise for investment management. As with most quantitative applications in finance, the danger of misapplying these techniques can lead to disappointment. One crucial limitation involves data availability. Many of machine learning early successes originated in the physical and biological sciences, in which truly vast amounts of data are available. Machine learning applications often require far more data than are available in finance, which is of particular concern in longer-horizon investing. Hence, choosing the right applications before applying the tools is important. In addition, capital markets reflect the actions of people, which may be influenced by others actions and by the findings of past research. In many ways, the challenges that affect machine learning are merely a continuation of the long-standing issues researchers have always faced in quantitative finance. While investors need to be cautious, more cautious than in past applications of quantitative methods new tools offer many potential applications in finance. In this article, the authors develop a research protocol that pertains both to the application of machine learning techniques and to quantitative finance in general.

Arora, S. and Doshi, P. (2018). "A Survey of Inverse Reinforcement Learning: Challenges, Methods and Progress." In: *arXiv e-Print*.
Inverse reinforcement learning is the problem of inferring the reward function of an observed agent, given its policy or behavior. Researchers perceive IRL both as a problem and as a class of methods. By categorically surveying the current literature in IRL, this article serves as a reference for researchers and practitioners in machine learning to understand the challenges of IRL and select the approaches best suited for the problem on hand. The survey formally introduces the IRL problem along with its central challenges which include accurate inference, generalizability, correctness of prior knowledge, and growth in solution complexity with problem size. The article elaborates how the current methods mitigate these challenges. We further discuss the extensions of traditional IRL methods: (i) inaccurate and incomplete perception, (ii) incomplete model, (iii) multiple rewards, and (iv) non-linear reward functions. This discussion concludes with some broad advances in the research area and currently open research questions.

Arulkumaran, K., Deisenroth, M. P., Brundage, M., and Bharath, A. A. (2017). "Deep reinforcement learning: A brief survey." In: *IEEE Signal Processing Magazine* 34(6), pp. 26–38.
Deep reinforcement learning (DRL) is poised to revolutionize the field of artificial intelligence (AI) and represents a step toward building autonomous systems with a higherlevel understanding of the visual world. Currently, deep learning is enabling reinforcement learning (RL) to scale to problems that were previously intractable, such as learning to play video games directly from pixels. DRL algorithms are also applied to robotics, allowing control policies for robots to be learned directly from camera inputs in the real world. In this survey, we begin with an introduction to the general field of RL, then progress to the main streams of value-based and policy-based methods. Our survey will cover central algorithms in deep RL, including the deep Q-network (DQN), trust region policy optimization (TRPO), and asynchronous advantage actor critic. In parallel, we highlight the unique advantages of deep neural networks, focusing on visual understanding via RL. To conclude, we describe several current areas of research within the field.

Bailey, D. H., Borwein, J. M., and Lopez de Prado, M. (2017). "Stock Portfolio Design and Backtest Overfitting." In: *Journal of Investment Management* 15(1), pp. 75–87.
In mathematical finance, backtest overfitting connotes the usage of historical market data to develop an investment strategy, where too many variations of the strategy are tried, relative to the amount of data available. Backtest overfitting is now thought to be a primary reason why investment models and strategies that look good on paper often disappoint in practice. Models and strategies suffering from overfitting typically target the specific idiosyncrasies of a limited dataset, rather than any general behavior, and, as a result, often perform erratically when presented with new data. In this study, we address overfitting in the context of designing a mutual fund or investment portfolio as a weighted collection of stocks. Very often a newly minted equity-based fund of this type has been designed by an exhaustive computer-based search of some sort to obtain an optimal weighting that exhibits excellent performance based, say, on the past 10 or 20 years' historical market data, and the fund often highlights this backtest performance.

Bareinboim, E. (2020). *Causal Reinforcement Learning*. URL: https://crl.causalai.net/.

This page provides information and materials about "Causal Reinforcement Learning " (CRL), following the tutorial presented at ICML 2020. For the corresponding material, check out the links Causal inference provides a set of tools and principles that allows one to combine data and structural invariances about the environment to reason about questions of counterfactual nature - i.e., what would have happened had reality been different, even when no data about this imagined reality is available. Reinforcement Learning is concerned with efficiently finding a policy that optimizes a specific function (e.g., reward, regret) in interactive and uncertain environments. These two disciplines have evolved independently and with virtually no interaction between them. In reality, however, they operate over different aspects of the same building block, i.e., counterfactual relations, which makes them umbilically tied. In this tutorial, we introduce a unified treatment based on this observation, putting these two disciplines under the same conceptual and theoretical umbrella. We show that a number of natural and pervasive classes of learning problems emerge when this connection is fully established, which cannot be seen individually from either discipline alone. In particular, we'll discuss generalized policy learning (a combination of online, off-policy, and do-calculus learning), when and where to intervene, counterfactual decision-making (and free-will, autonomy, human-AI collaboration), policy generalizability, and causal imitation learning, among others. This new understanding leads to a broader view of what counterfactual learning is, and suggests the great potential for the study of causality and reinforcement learning side by side. We call this new line of investigation "Causal Reinforcement Learning" (CRL, for short).

Bargiacchi, E., Roijers, D. M., and Nowe, A. (2020). "AI-Toolbox: A C++ library for Reinforcement Learning and Planning (with Python Bindings)." In: *Journal of Machine Learning Research*.
This paper describes AI-Toolbox, a C++ software library that contains reinforcement learning and planning algorithms, and supports both single and multi agent problems, as well as partial observability. It is designed for simplicity and clarity, and contains extensive documentation of its API and code. It supports Python to enable users not comfortable with C++ to take advantage of the library's speed and functionality. AI-Toolbox is free software, and is hosted online at `https://github.com/Svalorzen/{AI}-Toolbox`.

Bartram, S. M., Branke, J., De Rossi, G., and Motahari, M. (2021). "Machine Learning for Active Portfolio Management." In: *The Journal of Financial Data Science* 3(3), pp. 9–30.
Machine learning (ML) methods are attracting considerable attention among academics in the field of finance. However, it is commonly believed that ML has not transformed the asset management industry to the same extent as other sectors. This survey focuses on the ML methods and empirical results available in the literature that matter most for active portfolio management. ML has asset management applications for signal generation, portfolio construction, and trade execution, and promising findings have been reported. Reinforcement learning (RL), in particular, is expected to play a more significant role in the industry. Nevertheless, the performance of a sample of active exchange-traded funds (ETF) that use ML in their investments tends to be mixed. Overall, ML techniques show great promise for active portfolio management, but investors should be cautioned against their main potential pitfalls.

Benhamou, E. (2021). "Next Generation Robo Advisors." In: *SSRN e-Print*.
The takeover of robots in the traditional field of Asset Management is an emerging trend across the industry. It has been growing much faster with the current digitalization wave. The purpose of this AI Dauphine Summer School class is to present the subject and gives an intuition of the next Robo-advisors generation powered by AI.

Benhamou, E., Saltiel, D., Ohana, J.-J., and Atif, J. (2021a). "Detecting and Adapting to Crisis Pattern with Context Based Deep Reinforcement Learning." In: *SSRN e-Print*.
Deep reinforcement learning (DRL) has reached super human levels in complex tasks like game solving (Go, StarCraft II, Atari Games), and autonomous driving. However, it remains an open question whether DRL can reach human level in applications to financial problems and in particular in detecting pattern crisis and consequently dis-investing. In this paper, we present an innovative DRL framework consisting in two subnetworks fed respectively with portfolio strategies past performances and standard deviations as well as additional contextual features. The second sub network plays an important role as it captures dependencies with common financial indicators features like risk aversion, economic surprise index and correlations between assets that allows taking into account context based information. We compare different network architectures either using layers of convolutions to reduce network's complexity or LSTM block to capture time dependency and whether previous allocations is important in the modeling. We also use adversarial training to make the final model more robust. Results on test set show this approach substantially over-performs traditional

portfolio optimization methods like Markovitz and is able to detect and anticipate crisis like the current COVID one.

Benhamou, E., Saltiel, D., Ohana, J.-J., Atif, J., and Laraki, R. (2021b). "Deep Reinforcement Learning (DRL) for Portfolio Allocation." In: *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pp. 527–531.

Deep reinforcement learning (DRL) has reached an unprecedent level on complex tasks like game solving (Go or StarCraft II), and autonomous driving. However, applications to real financial assets are still largely unexplored and it remains an open question whether DRL can reach super human level. In this ECML PKKDD demo, we showcase state-of-the-art DRL methods for selecting portfolios according to financial environment, with a final network concatenating three individual networks using layers of convolutions to reduce network's complexity.

The multi entries of our network enables capturing dependencies from common financial indicators features like risk aversion, citigroup index surprise, portfolio specific features and previous portfolio allocations. Results on test set show this approach can overperform traditional portfolio optimization methods, with results available at our demo website `http://www.aisquareconnect.com/deeprl/models.html`.

Benhamou, E., Saltiel, D., Ohana, J.-J., Atif, J., and Laraki, R. (2021c). "Deep Reinforcement Learning (DRL) for Portfolio Allocation." In: *SSRN e-Print*.

Deep reinforcement learning (DRL) has reached an unprecedent level on complex tasks like game solving (Go or StarCraft II), and autonomous driving. However, applications to real financial assets are still largely unexplored and it remains an open question whether DRL can reach super human level. In this ECML PKKDD demo, we showcase state-of-the-art DRL methods for selecting portfolios according to financial environment, with a final network concatenating three individual networks using layers of convolutions to reduce network's complexity. The multi entries of our network enables capturing dependencies from common financial indicators features like risk aversion, citigroup index surprise, portfolio specific features and previous portfolio allocations. Results on test set show this approach can overperform traditional portfolio optimization methods.

Benhamou, E., Saltiel, D., Ohana, J.-J., Atif, J., and Laraki, R. (2021d). "DRLPS: Deep Reinforcement Learning for Portfolio Selection (Presentation Slides ECML PKDD)." In: *SSRN e-Print*.

Deep reinforcement learning (DRL) has reached an unprecedent level on complex tasks like game solving (Go or StarCraft II), and autonomous driving. However, applications to real financial assets are still largely unexplored and it remains an open question whether DRL can reach super human level. In this presentation for the ECML PKDD demo track, we showcase state-of-the-art DRL methods for selecting portfolios according to financial environment, with a final network concatenating three individual networks using layers of convolutions to reduce network's complexity. The multi entries of our network enables capturing dependencies from common financial indicators features like risk aversion, citigroup index surprise, portfolio specific features and previous portfolio allocations. Results on test set show this approach can overperform traditional portfolio optimization methods.

Benhamou, E., Saltiel, D., Tabachnik, S., Wong, S. K., and Chareyron, F. (2021e). "Adaptive learning for financial markets mixing model-based and model-free RL for volatility targeting." In: *arXiv e-Print*.

Model-Free Reinforcement Learning has achieved meaningful results in stable environments but, to this day, it remains problematic in regime changing environments like financial markets. In contrast, model-based RL is able to capture some fundamental and dynamical concepts of the environment but suffer from cognitive bias. In this work, we propose to combine the best of the two techniques by selecting various model-based approaches thanks to Model-Free Deep Reinforcement Learning. Using not only past performance and volatility, we include additional contextual information such as macro and risk appetite signals to account for implicit regime changes. We also adapt traditional RL methods to real-life situations by considering only past data for the training sets. Hence, we cannot use future information in our training data set as implied by K-fold cross validation. Building on traditional statistical methods, we use the traditional "walk-forward analysis", which is defined by successive training and testing based on expanding periods, to assert the robustness of the resulting agent. Finally, we present the concept of statistical difference's significance based on a two-tailed T-test, to highlight the ways in which our models differ from more traditional ones. Our experimental results show that our approach outperforms traditional financial baseline portfolio models such as the Markowitz model in almost all evaluation metrics commonly used in financial mathematics, namely net performance, Sharpe and Sortino ratios, maximum drawdown, maximum drawdown over volatility.

Benhamou, E., Saltiel, D., Tabachnik, S., Wong, S. K., and Chareyron, F. (2021f). "Adaptive learning for financial markets mixing model-based and model-free RL for volatility targeting." In: *SSRN e-Print*.
Model-Free Reinforcement Learning has achieved meaningful results in stable environments but, to this day, it remains problematic in regime changing environments like financial markets. In contrast, model-based RL is able to capture some fundamental and dynamical concepts of the environment but suffer from cognitive bias. In this work, we propose to combine the best of the two techniques by selecting various model-based approaches thanks to Model-Free Deep Reinforcement Learning. Using not only past performance and volatility, we include additional contextual information such as macro and risk appetite signals to account for implicit regime changes. We also adapt traditional RL methods to real-life situations by considering only past data for the training sets. Hence, we cannot use future information in our training data set as implied by K-fold cross validation. Building on traditional statistical methods, we use the traditional "walk-forward analysis", which is defined by successive training and testing based on expanding periods, to assert the robustness of the resulting agent. Finally, we present the concept of statistical difference's significance based on a two-tailed T-test, to highlight the ways in which our models differ from more traditional ones. Our experimental results show that our approach outperforms traditional financial baseline portfolio models such as the Markowitz model in almost all evaluation metrics commonly used in financial mathematics, namely net performance, Sharpe and Sortino ratios, maximum drawdown, maximum drawdown over volatility.

Benhamou, E., Saltiel, D., Ungari, S., Atif, J., and Mukhopadhyay, A. (2021g). "AAMDRL: Augmented Asset Management With Deep Reinforcement Learning." In: *SSRN e-Print*.
Can an agent learn efficiently in a noisy and self adapting environment with sequential, non-stationary and non-homogeneous observations? Through trading bots, we illustrate how Deep Reinforcement Learning (DRL) can tackle this challenge. Our contributions are threefold: (i) the use of contextual information also referred to as augmented state in DRL, (ii) the impact of a one period lag between observations and actions that is more realistic for an asset management environment, (iii) the implementation of a new repetitive train test method called walk forward analysis, similar in spirit to cross validation for time series. Although our experiment is on trading bots, it can easily be translated to other bot environments that operate in sequential environment with regime changes and noisy data. Our experiment for an augmented asset manager interested in finding the best portfolio for hedging strategies shows that AAMDRL achieves superior returns and lower risk.

Benhamou, E., Saltiel, D., Ungari, S., and Mukhopadhyay, A. (2020). "Time your hedge with Deep Reinforcement Learning." In: *arXiv e-Print*.
Can an asset manager plan the optimal timing for her/his hedging strategies given market conditions? The standard approach based on Markowitz or other more or less sophisticated financial rules aims to find the best portfolio allocation thanks to forecasted expected returns and risk but fails to fully relate market conditions to hedging strategies decision. In contrast, Deep Reinforcement Learning (DRL) can tackle this challenge by creating a dynamic dependency between market information and hedging strategies allocation decisions. In this paper, we present a realistic and augmented DRL framework that: (i) uses additional contextual information to decide an action, (ii) has a one period lag between observations and actions to account for one day lag turnover of common asset managers to rebalance their hedge, (iii) is fully tested in terms of stability and robustness thanks to a repetitive train test method called anchored walk forward training, similar in spirit to k fold cross validation for time series and (iv) allows managing leverage of our hedging strategy. Our experiment for an augmented asset manager interested in sizing and timing his hedges shows that our approach achieves superior returns and lower risk.

Benhamou, E., Saltiel, D., Ungari, S., and Mukhopadhyay, A. (2021h). "Bridging the Gap Between Markowitz Planning and Deep Reinforcement Learning." In: *SSRN e-Print*.
While researchers in the asset management industry have mostly focused on techniques based on financial and risk planning techniques like Markowitz efficient frontier, minimum variance, maximum diversification or equal risk parity, in parallel, another community in machine learning has started working on reinforcement learning and more particularly deep reinforcement learning to solve other decision making problems for challenging task like autonomous driving, robot learning, and on a more conceptual side games solving like Go. This paper aims to bridge the gap between these two approaches by showing Deep Reinforcement Learning (DRL) techniques can shed new lights on portfolio allocation thanks to a more general optimization setting that casts portfolio allocation as an optimal control problem that is not just a one-step optimization, but rather a continuous control optimization with a delayed reward. The advantages are numerous: (i) DRL

maps directly market conditions to actions by design and hence should adapt to changing environment, (ii) DRL does not rely on any traditional financial risk assumptions like that risk is represented by variance, (iii) DRL can incorporate additional data and be a multi inputs method as opposed to more traditional optimization methods. We present on an experiment some encouraging results using convolution networks.

Benhamou, E., Saltiel, D., Ungari, S., and Mukhopadhyay, A. (2021i). "Bridging the gap between Markowitz planning and deep reinforcement learning (ICAPS PRL Presentation Slides 2020)." In: *SSRN e-Print*.

While researchers in the asset management industry have mostly focused on techniques based on financial and risk planning techniques like Markowitz efficient frontier, minimum variance, maximum diversification or equal risk parity, in parallel, another community in machine learning has started working on reinforcement learning and more particularly deep reinforcement learning to solve other decision making problems for challenging task like autonomous driving, robot learning, and on a more conceptual side games solving like Go.

This paper aims to bridge the gap between these two approaches by showing Deep Reinforcement Learning (DRL) techniques can shed new lights on portfolio allocation thanks to a more general optimization setting that casts portfolio allocation as an optimal control problem that is not just a one-step optimization, but rather a continuous control optimization with a delayed reward. The advantages are numerous: (i) DRL maps directly market conditions to actions by design and hence should adapt to changing environment, (ii) DRL does not rely on any traditional financial risk assumptions like that risk is represented by variance, (iii) DRL can incorporate additional data and be a multi inputs method as opposed to more traditional optimization methods. We present on an experiment some encouraging results using convolution networks.

Bertsekas, D. P. (2018). "Feature-Based Aggregation and Deep Reinforcement Learning: A Survey and Some New Implementations." In: *arXiv e-Print*.

In this paper we discuss policy iteration methods for approximate solution of a finite-state discounted Markov decision problem, with a focus on feature-based aggregation methods and their connection with deep reinforcement learning schemes. We introduce features of the states of the original problem, and we formulate a smaller "aggregate" Markov decision problem, whose states relate to the features. We discuss properties and possible implementations of this type of aggregation, including a new approach to approximate policy iteration. In this approach the policy improvement operation combines feature-based aggregation with feature construction using deep neural networks or other calculations. We argue that the cost function of a policy may be approximated much more accurately by the nonlinear function of the features provided by aggregation, than by the linear function of the features provided by neural network-based reinforcement learning, thereby potentially leading to more effective policy improvement.

Bessler, W., Opfer, H., and Wolff, D. (2017). "Multi-asset portfolio optimization and out-of-sample performance: an evaluation of Black Litterman, mean-variance, and naive diversification approaches." In: *The European Journal of Finance* 23(1), pp. 1–30.

The Black Litterman model aims to enhance asset allocation decisions by overcoming the problems of mean-variance portfolio optimization. We propose a sample-based version of the Black Litterman model and implement it on a multi-asset portfolio consisting of global stocks, bonds, and commodity indices, covering the period from January 1993 to December 2011. We test its out-of-sample performance relative to other asset allocation models and find that Black Litterman optimized portfolios significantly outperform na ive-diversified portfolios (1/N rule and strategic weights), and consistently perform better than mean-variance, Bayes Stein, and minimum-variance strategies in terms of out-of-sample Sharpe ratios, even after controlling for different levels of risk aversion, investment constraints, and transaction costs. The BL model generates portfolios with lower risk, less extreme asset allocations, and higher diversification across asset classes. Sensitivity analyses indicate that these advantages are due to more stable mixed return estimates that incorporate the reliability of return predictions, smaller estimation errors, and lower turnover.

Bessler, W. and Wolff, D. (2017). "Portfolio Optimization with Industry Return Prediction Models." In: *SSRN e-Print*.

We postulate that utilizing return prediction models with fundamental, macroeconomic, and technical indicators instead of using historical averages should result in superior asset allocation decisions. We investigate the predictive power of individual variables for forecasting industry returns in-sample and out-of-sample and then analyze multivariate predictive regression models including OLS, a regularization technique, principal components, a target-relevant latent factor approach, and forecast combinations. The gains from using industry return predictions are evaluated in an out-of-sample Black-Litterman portfolio optimization frame-

work. We provide empirical evidence that portfolio optimization utilizing industry return prediction models significantly outperform portfolios using historical averages and those being passively managed.

Beysolow II, T. (2019). *Applied Reinforcement Learning with Python: With OpenAI Gym, Tensorflow, and Keras*. Berkeley, CA: Apress. 112 pp.
Delve into the world of reinforcement learning algorithms and apply them to different use-cases via Python. This book covers important topics such as policy gradients and Q learning, and utilizes frameworks such as Tensorflow, Keras, and OpenAI Gym. Applied Reinforcement Learning with Python introduces you to the theory behind reinforcement learning (RL) algorithms and the code that will be used to implement them. You will take a guided tour through features of OpenAI Gym, from utilizing standard libraries to creating your own environments, then discover how to frame reinforcement learning problems so you can research, develop, and deploy RL-based solutions.

Bjerring, T., Ross, O., and Weissensteiner, A. (2017). "Feature selection for portfolio optimization." In: *Annals of Operations Research* 256, pp. 21–40.
Most portfolio selection rules based on the sample mean and covariance matrix perform poorly out-of-sample. Moreover, there is a growing body of evidence that such optimization rules are not able to beat simple rules of thumb, such as 1/N. Parameter uncertainty has been identified as one major reason for these findings. A strand of literature addresses this problem by improving the parameter estimation and/or by relying on more robust portfolio selection methods. Independent of the chosen portfolio selection rule, we propose using feature selection first in order to reduce the asset menu. While most of the diversification benefits are preserved, the parameter estimation problem is alleviated. We conduct out-of-sample back-tests to show that in most cases different well-established portfolio selection rules applied on the reduced asset universe are able to improve alpha relative to different prominent factor models.

Borrageiro, G., Firoozye, N., and Barucca, P. (2021). "Reinforcement Learning for Systematic FX Trading." In: *arXiv e-Print*.
We conduct a detailed experiment on major cash fx pairs, accurately accounting for transaction and funding costs. These sources of profit and loss, including the price trends that occur in the currency markets, are made available to our recurrent reinforcement learner via a quadratic utility, which learns to target a position directly. We improve upon earlier work, by casting the problem of learning to target a risk position, in an online learning context. This online learning occurs sequentially in time, but also in the form of transfer learning. We transfer the output of radial basis function hidden processing units, whose means, covariances and overall size are determined by Gaussian mixture models, to the recurrent reinforcement learner and baseline momentum trader. Thus the intrinsic nature of the feature space is learnt and made available to the upstream models. The recurrent reinforcement learning trader achieves an annualised portfolio information ratio of 0.52 with compound return of 9.3%, net of execution and funding cost, over a 7 year test set. This is despite forcing the model to trade at the close of the trading day 5pm EST, when trading costs are statistically the most expensive. These results are comparable with the momentum baseline trader, reflecting the low interest differential environment since the the 2008 financial crisis, and very obvious currency trends since then. The recurrent reinforcement learner does nevertheless maintain an important advantage, in that the model's weights can be adapted to reflect the different sources of profit and loss variation. This is demonstrated visually by a USDRUB trading agent, who learns to target different positions, that reflect trading in the absence or presence of cost.

Borrageiro, G., Firoozye, N., and Barucca, P. (2022). "The Recurrent Reinforcement Learning Crypto Agent." In: *arXiv e-Print*.
We demonstrate an application of online transfer learning as a digital assets trading agent. This agent makes use of a powerful feature space representation in the form of an echo state network, the output of which is made available to a direct, recurrent reinforcement learning agent. The agent learns to trade the XBTUSD (Bitcoin versus US dollars) perpetual swap derivatives contract on BitMEX. It learns to trade intraday on five minutely sampled data, avoids excessive over-trading, captures a funding profit and is also able to predict the direction of the market. Overall, our crypto agent realises a total return of 350%, net of transaction costs, over roughly five years, 71% of which is down to funding profit. The annualised information ratio that it achieves is 1.46.

Bou, A. and De Fabritiis, G. (2021). "PyTorchRL: Modular and Distributed Reinforcement Learning in PyTorch." In: *arXiv e-Print*.

Deep reinforcement learning (RL) has proved successful at solving challenging environments but often requires scaling to large sampling and computing resources. Furthermore, advancing RL requires tools that are flexible enough to easily prototype new methods, yet avoiding impractically slow experimental turnaround times. To this end, we present PyTorchRL, a PyTorch-based library for RL with a modular design that allows composing agents from a set of reusable and easily extendable modules. Additionally, PyTorchRL permits the definition of distributed training architectures with flexibility and independence of the Agent components. In combination, these two features can accelerate the pace at which ideas are implemented and tested, simplifying research and enabling to tackle more challenging RL problems. We present several interesting use-cases of PyTorchRL and showcase the library by obtaining the highest to-date test performance on the Obstacle Tower Unity3D challenge environment.

Brini, A. and Tantari, D. (2022). "Deep Reinforcement Trading with Predictable Returns." In: *arXiv e-Print*.
Classical portfolio optimization often requires forecasting asset returns and their corresponding variances in spite of the low signal-to-noise ratio provided in the financial markets. Modern deep reinforcement learning (DRL) offers a framework for optimizing sequential trader decisions but lacks theoretical guarantees of convergence. On the other hand the performances on real financial trading problems are strongly affected by the goodness of the signal used to predict returns. To disentangle the effects coming from return unpredictability from those coming from algorithm un-trainability, we investigate the performance of model-free DRL traders in a market environment with different known mean-reverting factors driving the dynamics. When the framework admits an exact dynamic programming solution, we can assess limits and capabilities of different value-based algorithms to retrieve meaningful trading signals in a data-driven manner. We consider DRL agents that leverage on classical strategies to increase their performances and we show that this approach guarantees flexibility, outperforming the benchmark strategy when the price dynamics is misspecified and some original assumptions on the market environment are violated with the presence of extreme events and volatility clustering.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W. (2016). "OpenAI Gym." In: *arXiv e-Print*.
OpenAI Gym is a toolkit for reinforcement learning research. It includes a growing collection of benchmark problems that expose a common interface, and a website where people can share their results and compare the performance of algorithms. This whitepaper discusses the components of OpenAI Gym and the design decisions that went into the software.

Brown, D. S., Cui, Y., and Niekum, S. (2019). "Risk-Aware Active Inverse Reinforcement Learning." In: *arXiv e-Print*.
Active learning from demonstration allows a robot to query a human for specific types of input to achieve efficient learning. Existing work has explored a variety of active query strategies; however, to our knowledge, none of these strategies directly minimize the performance risk of the policy the robot is learning. Utilizing recent advances in performance bounds for inverse reinforcement learning, we propose a risk-aware active inverse reinforcement learning algorithm that focuses active queries on areas of the state space with the potential for large generalization error. We show that risk-aware active learning outperforms standard active IRL approaches on gridworld, simulated driving, and table setting tasks, while also providing a performance-based stopping criterion that allows a robot to know when it has received enough demonstrations to safely perform a task.

Brown, D. S. and Niekum, S. (2019). "Machine teaching for inverse reinforcement learning: algorithms and applications." In: *Proceedings of the AAAI Conference on Artificial Intelligence* 33, pp. 7749–7758.
Inverse reinforcement learning (IRL) infers a reward function from demonstrations, allowing for policy improvement and generalization. However, despite much recent interest in IRL, little work has been done to understand the minimum set of demonstrations needed to teach a specific sequential decisionmaking task. We formalize the problem of finding maximally informative demonstrations for IRL as a machine teaching problem where the goal is to find the minimum number of demonstrations needed to specify the reward equivalence class of the demonstrator. We extend previous work on algorithmic teaching for sequential decision-making tasks by showing a reduction to the set cover problem which enables an efficient approximation algorithm for determining the set of maximallyinformative demonstrations. We apply our proposed machine teaching algorithm to two novel applications: providing a lower bound on the number of queries needed to learn a policy using active IRL and developing a novel IRL algorithm that can learn more efficiently from informative demonstrations than a standard IRL approach.

Brunel, J. L. P. (2020). "Extending the Goals-Based Framework to Comprise Both Investment and Financial Planning." In: *The Journal of Wealth Management* 22(44), pp. 21–26.
This article seeks to fill a void in the literature on goals-based planning. Most of the current work covers cases where clients already possess significant financial assets that they plan on totally or partially spending down, through expenses as well as various dynastic or philanthropic transfers. Yet, planners–be they focused on income/savings/expense management issues or conduct their work in the asset management sphere–have to deal with at least two other potential cases. Our analysis suggests that the goals-based planning approach has the potential to inject more texture into conversations with all clients. It also shows that there are significant, and at times even dramatic, differences in the ways one might deal with 1) an individual who is already wealthy and spending down his or her wealth, 2) another who may be wealthy, but has unrealized non-financial wealth and ongoing current savings inflows, and 3) yet another who has a significant income and savings power but not enough accumulated financial wealth to live on. This is nothing more than the classical case of the difference between human and financial capital. Individuals in an asset decumulation mode have more financial than human capital, while those who are in an accumulation mode have more human than financial capital. As one might expect, this describes a full spectrum, and our "accumulation/decumulation case" falls somewhere between these extremes.

Bruni, R., Cesarone, F., Scozzari, A., and Tardella, F. (2016). "Real-world datasets for portfolio selection and solutions of some stochastic dominance portfolio models." In: *Data in Brief* 8, pp. 858–862.
A large number of portfolio selection models have appeared in the literature since the pioneering work of Markowitz. However, even when computational and empirical results are described, they are often hard to replicate and compare due to the unavailability of the datasets used in the experiments. We provide here several datasets for portfolio selection generated using real-world price values from several major stock markets. The datasets contain weekly return values, adjusted for dividends and for stock splits, which are cleaned from errors as much as possible. The datasets are available in different formats, and can be used as benchmarks for testing the performances of portfolio selection models and for comparing the efficiency of the algorithms used to solve them. We also provide, for these datasets, the portfolios obtained by several selection strategies based on Stochastic Dominance models (see "On Exact and Approximate Stochastic Dominance Strategies for Portfolio Selection" (Bruni et al. [2])). We believe that testing portfolio models on publicly available datasets greatly simplifies the comparison of the different portfolio selection strategies.

Bruni, R., Cesarone, F., Scozzari, A., and Tardella, F. (2017). "On exact and approximate stochastic dominance strategies for portfolio selection." In: *European Journal of Operational Research* 259(1), pp. 322–329.
New type of approximate stochastic dominance designed for portfolio selection. Equivalent to minimizing the expected shortfall of the portfolio below the benchmark. An easily solvable LP model for the practical implementation of our approach. Extensive empirical comparison of stochastic dominance models for portfolio selection. One recent and promising strategy for Enhanced Indexation is the selection of portfolios that stochastically dominate the benchmark. We propose here a new type of approximate stochastic dominance rule which implies other existing approximate stochastic dominance rules. We then use it to find the portfolio that approximately stochastically dominates a given benchmark with the best possible approximation. Our model is initially formulated as a Linear Program with exponentially many constraints, and then reformulated in a more compact manner so that it can be very efficiently solved in practice. This reformulation also reveals an interesting financial interpretation. We compare our approach with several exact and approximate stochastic dominance models for portfolio selection. An extensive empirical analysis on real and publicly available datasets shows very good out-of-sample performances of our model.

Bryzgalova, S., Huang, J., and Julliard, C. (2021). "Bayesian solutions for the factor zoo: we just ran two quadrillion models." In: *SSRN e-Print*.
We propose a novel, and simple, Bayesian estimation and model selection procedure for cross-sectional asset pricing. Our approach, that allows for both tradable and non-tradable factors, and is applicable to high dimensional cases, has several desirable properties. First, weak and spurious factors lead to diffuse, and centered at zero, posteriors for their market price of risk, making such factors easily detectable. Second, posterior inference is robust to the presence of such factors. Third, we show that flat priors for risk premia lead to improper marginal likelihoods, rendering model selection invalid. Therefore, we provide a novel prior, that is diffuse for strong factors but shrinks away useless ones, under which posterior probabilities are well behaved, and can be used for factor and (non necessarily nested) model selection, as well as model averaging,

in large scale problems. We apply our method to a very large set of factors proposed in the literature, and analyse 2.25 quadrillion possible models, gaining novel insights on the empirical drivers of asset returns.

Buehler, H., Gonon, L., Teichmann, J., and Wood, B. (2019). "Deep hedging." In: *Quantitative Finance* 19 (8), pp. 1271–1291.
We present a framework for hedging a portfolio of derivatives in the presence of market frictions such as transaction costs, liquidity constraints or risk limits using modern deep reinforcement machine learning methods. We discuss how standard reinforcement learning methods can be applied to non-linear reward structures, i.e. in our case convex risk measures. As a general contribution to the use of deep learning for stochastic processes, we also show in Section 4 that the set of constrained trading strategies used by our algorithm is large enough to eps-approximate any optimal solution. Our algorithm can be implemented efficiently even in high-dimensional situations using modern machine learning tools. Its structure does not depend on specific market dynamics, and generalizes across hedging instruments including the use of liquid derivatives. Its computational performance is largely invariant in the size of the portfolio as it depends mainly on the number of hedging instruments available. We illustrate our approach by an experiment on the SandP500 index and by showing the effect on hedging under transaction costs in a synthetic market driven by the Heston model, where we outperform the standard -market solution.

Cai, Q., Cui, C., Xiong, Y., Wang, W., Xie, Z., and Zhang, M. (2022). "A Survey on Deep Reinforcement Learning for Data Processing and Analytics." In: *arXiv e-Print*.
Data processing and analytics are fundamental and pervasive. Algorithms play a vital role in data processing and analytics where many algorithm designs have incorporated heuristics and general rules from human knowledge and experience to improve their effectiveness. Recently, reinforcement learning, deep reinforcement learning (DRL) in particular, is increasingly explored and exploited in many areas because it can learn better strategies in complicated environments it is interacting with than statically designed algorithms. Motivated by this trend, we provide a comprehensive review of recent works focusing on utilizing deep reinforcement learning to improve data processing and analytics. First, we present an introduction to key concepts, theories, and methods in deep reinforcement learning. Next, we discuss deep reinforcement learning deployment on database systems, facilitating data processing and analytics in various aspects, including data organization, scheduling, tuning, and indexing. Then, we survey the application of deep reinforcement learning in data processing and analytics, ranging from data preparation, natural language interface to healthcare, fintech, etc. Finally, we discuss important open challenges and future research directions of using deep reinforcement learning in data processing and analytics.

Cao, J., Chen, J., Hull, J., and Poulos, Z. (2021). "Deep Hedging of Derivatives Using Reinforcement Learning." In: *The Journal of Financial Data Science* 3(1), pp. 10–27.
This article shows how reinforcement learning can be used to derive optimal hedging strategies for derivatives when there are transaction costs. The article illustrates the approach by showing the difference between using delta hedging and optimal hedging for a short position in a call option when the objective is to minimize a function equal to the mean hedging cost plus a constant times the standard deviation of the hedging cost. Two situations are considered. In the first, the asset price follows a geometric Brownian motion. In the second, the asset price follows a stochastic volatility process. The article extends the basic reinforcement learning approach in several ways. First, it uses two different Q-functions to track both the expected value of the cost and the expected value of the square of the cost. This approach increases the range of objective functions that can be used. Second, it uses a learning algorithm that allows for continuous state and action space. Third, it compares the accounting profit and loss (P&L) approach and the cash flow approach. The authors find that a hybrid approach involving the use of an accounting P&L approach that incorporates a relatively simple valuation model works well.

Cao, J., Chen, J., Hull, J. C., and Poulos, Z. (2020). "Deep Hedging of Derivatives Using Reinforcement Learning." In: *SSRN e-Print*.
This paper shows how reinforcement learning can be used to derive optimal hedging strategies for derivatives when there are transaction costs. The paper illustrates the approach by showing the difference between using delta hedging and optimal hedging for a short position in a call option when the objective is to minimize a function equal to the mean hedging cost plus a constant times the standard deviation of the hedging cost. Two situations are considered. In the first, the asset price follows geometric Brownian motion. In the second, the asset price follows a stochastic volatility process. The paper extends the basic reinforcement learning approach in a number of ways. First, it uses two different Q-functions so that both the expected value of the

cost and the expected value of the square of the cost are tracked for different state/action combinations. This approach increases the range of objective functions that can be used. Second, it uses a learning algorithm that allows for continuous state and action space. Third, it compares the accounting P&L approach (where the hedged position is valued at each step) and the cash flow approach (where cash infloows and outflows are used). We find that a hybrid approach involving the use of an accounting P&L approach that incorporates a relatively simple valuation model works well. The valuation model does not have to correspond to the process assumed for the underlying asset price.

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

This study presents a deep reinforcement learning approach for global hedging of long-term financial derivatives. A similar setup as in Coleman et al. (2007) is considered with the risk management of lookback options embedded in guarantees of variable annuities with ratchet features. The deep hedging algorithm of Buehler et al. (2019a) is applied to optimize neural networks representing global hedging policies with both quadratic and non-quadratic penalties. To the best of the author's knowledge, this is the first paper that presents an extensive benchmarking of global policies for long-term contingent claims with the use of various hedging instruments (e.g. underlying and standard options) and with the presence of jump risk for equity. Monte Carlo experiments demonstrate the vast superiority of non-quadratic global hedging as it results simultaneously in downside risk metrics two to three times smaller than best benchmarks and in significant hedging gains. Analyses show that the neural networks are able to effectively adapt their hedging decisions to different penalties and stylized facts of risky asset dynamics only by experiencing simulations of the financial market exhibiting these features. Numerical results also indicate that non-quadratic global policies are significantly more geared towards being long equity risk which entails earning the equity risk premium.

Carbonneau, A. and Godin, F. (2021). "Equal risk pricing of derivatives with deep hedging." In: *Quantitative Finance* 21(4), pp. 593–608.

This article provides a universal and tractable methodology based on deep reinforcement learning to implement the equal risk pricing framework for financial derivatives pricing under very general conditions. The equal risk pricing framework entails solving for a derivative price which equates the optimally hedged residual risk exposure associated, respectively, with the long and short positions in the contingent claim. The solution to the hedging optimization problem considered, which is inspired from the [Marzban, S., Delage, E. and Li, J.Y., Equal risk pricing and hedging of financial derivatives with convex risk measures. arXiv preprint arXiv:2002.02876, 2020.] framework relying on convex risk measures, is obtained through the use of the deep hedging algorithm of [Buehler, H., Gonon, L., Teichmann, J. and Wood, B., Deep hedging. Q. Finance, 2019, 19, 1271-1291]. Consequently, the current paper's approach allows for the pricing and the hedging of a very large number of contingent claims (e.g. vanilla options, exotic options, options with multiple underlying assets) with multiple liquid hedging instruments under a wide variety of market dynamics (e.g. regime-switching, stochastic volatility, jumps). A novel epsilon-completeness measure allowing for the quantification of the residual hedging risk associated with a derivative is also proposed. The latter measure generalizes the one presented in [Bertsimas, D., Kogan, L. and Lo, A.W., Hedging derivative securities and incomplete markets: an epsilon-arbitrage approach. Oper. Res., 2001, 49, 372-397.] based on the quadratic penalty. Monte Carlo simulations are performed under a large variety of market dynamics to demonstrate the practicability of our approach, to perform benchmarking with respect to traditional methods and to conduct sensitivity analyses. Numerical results show, among others, that equal risk prices of out-of-the-money options are significantly higher than risk-neutral prices stemming from conventional changes of measure across all dynamics considered. This finding is shown to be shared by different option categories which include vanilla and exotic options.

Cartea, A., Jaimungal, S., and Sanchez-Betancourt, L. (2022). "Deep Reinforcement Learning for Algorithmic Trading." In: *Machine Learning in Financial Markets: A guide to contemporary practices*. Ed. by A. Capponi and C.-A. Lehalle. Cambridge University Press.

We employ reinforcement learning (RL) techniques to devise statistical arbitrage strategies in electronic markets. In particular, double deep Q network learning (DDQN) and a new variant of reinforced deep Markov models (RDMMs) are used to derive the optimal strategies for an agent who trades in a foreign exchange (FX) triplet. An FX triplet consists of three currency pairs where the exchange rate of one pair is redundant because, by no-arbitrage, it is determined by the exchange rates of the other two pairs. We use simulations of a co-integrated model of exchange rates to implement the strategies and show their financial performance.

Castro, P. S., Li, S., and Zhang, D. (2019). "Inverse Reinforcement Learning with Multiple Ranked Experts." In: *arXiv e-Print*.
We consider the problem of learning to behave optimally in a Markov Decision Process when a reward function is not specified, but instead we have access to a set of demonstrators of varying performance. We assume the demonstrators are classified into one of k ranks, and use ideas from ordinal regression to find a reward function that maximizes the margin between the different ranks. This approach is based on the idea that agents should not only learn how to behave from experts, but also how not to behave from non-experts. We show there are MDPs where important differences in the reward function would be hidden from existing algorithms by the behaviour of the expert. Our method is particularly useful for problems where we have access to a large set of agent behaviours with varying degrees of expertise (such as through GPS or cellphones). We highlight the differences between our approach and existing methods using a simple grid domain and demonstrate its efficacy on determining passenger-finding strategies for taxi drivers, using a large dataset of GPS trajectories.

Castro, P. S., Moitra, S., Gelada, C., Kumar, S., and Bellemare, M. G. (2018). "Dopamine: A Research Framework for Deep Reinforcement Learning." In: *arXiv e-Print*.
Deep reinforcement learning (deep RL) research has grown significantly in recent years. A number of software offerings now exist that provide stable, comprehensive implementations for benchmarking. At the same time, recent deep RL research has become more diverse in its goals. In this paper we introduce Dopamine, a new research framework for deep RL that aims to support some of that diversity. Dopamine is open-source, TensorFlow-based, and provides compact and reliable implementations of some state-of-the-art deep RL agents. We complement this offering with a taxonomy of the different research objectives in deep RL research. While by no means exhaustive, our analysis highlights the heterogeneity of research in the field, and the value of frameworks such as ours.

Cesarone, F., Moretti, J., and Tardella, F. (2018). "Why Small Portfolios Are Preferable and How to Choose Them." In: *SSRN e-Print*.
One of the fundamental principles in portfolio selection models is minimization of risk through diversification of the investment. However, this principle does not necessarily translate into a request for investing in all the assets of the investment universe. Indeed, following a line of research started by Evans and Archer almost 50 years ago, we provide here further evidence that small portfolios are sufficient to achieve almost optimal in-sample risk reduction with respect to variance and to some other popular risk measures, and very good out-of-sample performances. While leading to similar results, our approach is significantly different from the classical one pioneered by Evans and Archer. Indeed, we describe models for choosing the portfolio of a prescribed size with the smallest possible risk, as opposed to the random portfolio choice investigated in most of the previous works. We find that the smallest risk portfolios generally require no more than 15 assets. Furthermore, it is almost always possible to find portfolios that are just 1% more risky than the smallest risk portfolios and contain no more than 10 assets. The preference for small optimal portfolios is also justified by recent theoretical results on the estimation errors for the parameters required by portfolio selection models. Our empirical analysis is based on some new and on some publicly available benchmark data sets often used in the literature.

Cesarone, F., Mottura, C., Ricci, J. M., and Tardella, F. (2019). "On the stability of portfolio selection models." In: *SSRN e-Print*.
One of the main issues in portfolio selection models consists in assessing the effect of the estimation errors of the parameters required by the models on the quality of the selected portfolios. Several studies have been devoted to this topic for the minimum variance and for several other minimum risk models. However, no sensitivity analysis seems to have been reported for the recent popular Risk Parity diversification approach, nor for other portfolio selection models requiring maximum gain-risk ratios.Based on artificial and real-world data, we provide here empirical evidence showing that the Risk Parity model is always the most stable one in all the cases analyzed. Furthermore, the minimum risk models are typically more stable than the maximum gain-risk models, with the minimum variance model often being the preferable one.

Chakraborty, S. (2019). "Deep Reinforcement Learning in Financial Markets." In: *arXiv e-Print*.
In this paper we explore the usage of deep reinforcement learning algorithms to automatically generate consistently profitable, robust, uncorrelated trading signals in any general financial market. In order to do this, we present a novel Markov decision process (MDP) model to capture the financial trading markets. We review and propose various modifications to existing approaches and explore different techniques to succinctly

capture the market dynamics to model the markets. We then go on to use deep reinforcement learning to enable the agent (the algorithm) to learn how to take profitable trades in any market on its own, while suggesting various methodology changes and leveraging the unique representation of the FMDP (financial MDP) to tackle the primary challenges faced in similar works. Through our experimentation results, we go on to show that our model could be easily extended to two very different financial markets and generates a positively robust performance in all conducted experiments.

Chan, Y., Hogan, K., Schwaiger, K., and Ang, A. (2020). "ESG in Factors." In: *The Journal of Impact and ESG Investing* 1(1), pp. 26–45.

Environmental, social, and governance (ESG) signals are an important part of factor-based investing strategies as they can stem from the same economic rationales as general factor premiums. Because factors are broad and diversified, building portfolios by jointly optimizing factor exposures with ESG and carbon outcomes can result in similar historical performance as benchmark factor portfolios that do not include those considerations. We show how sustainable signals, which often involve alternative data, can be integrated in the definitions of factors themselves: We offer two examples on green intangible value and corporate culture quality which enhance traditional financial value and quality factors, respectively.

Charpentier, A., Elie, R., and Remlinger, C. (2020). "Reinforcement Learning in Economics and Finance." In: *arXiv e-Print*.

Reinforcement learning algorithms describe how an agent can learn an optimal action policy in a sequential decision process, through repeated experience. In a given environment, the agent policy provides him some running and terminal rewards. As in online learning, the agent learns sequentially. As in multi-armed bandit problems, when an agent picks an action, he can not infer ex-post the rewards induced by other action choices. In reinforcement learning, his actions have consequences: they influence not only rewards, but also future states of the world. The goal of reinforcement learning is to find an optimal policy - a mapping from the states of the world to the set of actions, in order to maximize cumulative reward, which is a long term strategy. Exploring might be sub-optimal on a short-term horizon but could lead to optimal long-term ones. Many problems of optimal control, popular in economics for more than forty years, can be expressed in the reinforcement learning framework, and recent advances in computational science, provided in particular by deep learning algorithms, can be used by economists in order to solve complex behavioral problems. In this article, we propose a state-of-the-art of reinforcement learning techniques, and present applications in economics, game theory, operation research and finance.

Chaudhuri, S. E. and Lo, A. W. (2019). "Dynamic Alpha: A Spectral Decomposition of Investment Performance Across Time Horizons." In: *Management Science* 65(9), pp. 4440–4450.

The value added by an active investor is traditionally measured using alpha, tracking error, and the information ratio. However, these measures do not characterize the dynamic component of investor activity, nor do they consider the time horizons over which weights are changed. In this paper, we propose a technique to measure the value of active investment that captures both the static and dynamic contributions of an investment process. This dynamic alpha is based on the decomposition of a portfolio's expected return into its frequency components using spectral analysis. The result is a static component that measures the portion of a portfolio's expected return resulting from passive investments and security selection and a dynamic component that captures the manager's timing ability across a range of time horizons. Our framework can be universally applied to any portfolio and is a useful method for comparing the forecast power of different investment processes. Several analytical and empirical examples are provided to illustrate the practical relevance of this decomposition.

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

The Chief Investment Officer of Merrill Lynch Wealth Management explains why goals, not markets, should be the primary focus of your investment strategy-and offers a practical, innovative framework for making smarter choices about aligning your goals to your investment strategy. The Aspirational Investor is a practical, innovative approach to managing wealth based on key goals and the careful allocation of risks rather than responding to the whims of the financial markets. Chhabra introduces his "Wealth Allocation Framework," which accommodates the three seemingly incompatible objectives that must underpin every sound wealth management plan: the need for financial security in the face of known and unknowable risks; the need to maintain current living standards over time despite inflation; and the need to pursue aspirational goals for wealth creation.

Coache, A. and Jaimungal, S. (2022). "Reinforcement Learning with Dynamic Convex Risk Measures." In: *arXiv e-Print*.
We develop an approach for solving time-consistent risk-sensitive stochastic optimization problems using model-free reinforcement learning (RL). Specifically, we assume agents assess the risk of a sequence of random variables using dynamic convex risk measures. We employ a time-consistent dynamic programming principle to determine the value of a particular policy, and develop policy gradient update rules. We further develop an actor-critic style algorithm using neural networks to optimize over policies. Finally, we demonstrate the performance and flexibility of our approach by applying it to optimization problems in statistical arbitrage trading and obstacle avoidance robot control.

Cong, L., Tang, K., Wang, J., and Zhang, Y. (2022). "AlphaPortfolio: Direct Construction Through Deep Reinforcement Learning and Interpretable AI." In: *SSRN e-Print*.
We directly optimize the objectives of portfolio management via reinforcement learning—an alternative to conventional supervised-learning-based paradigms that entail first-step estimations of return distributions, pricing kernels, or risk premia. Building upon breakthroughs in AI, we develop multi-sequence neural network models tailored to distinguishing features of economic and financial data, while allowing training without labels and potential market interactions. The resulting AlphaPortfolio yields stellar out-of-sample performances (e.g., Sharpe ratio above two and over 13% risk-adjusted alpha with monthly re-balancing) that are robust under various economic restrictions and market conditions (e.g., exclusion of small stocks and short-selling). Moreover, we project AlphaPortfolio onto simpler modeling spaces (e.g., using polynomial-feature-sensitivity) to uncover key drivers of investment performance, including their rotation and nonlinearity. More generally, we highlight the utility of deep reinforcement learning in finance and invent "economic distillation" tools for interpreting AI and big data models.

Cong, L. W., Tang, K., Wang, J., and Zhang, Y. (2021). "Deep Sequence Modeling: Development and Applications in Asset Pricing." In: *The Journal of Financial Data Science* 3(1), pp. 28–42.
The authors predict asset returns and measure risk premiums using a prominent technique from artificial intelligence: deep sequence modeling. Because asset returns often exhibit sequential dependence that may not be effectively captured by conventional time-series models, sequence modeling offers a promising path with its data-driven approach and superior performance. In this article the authors first overview the development of deep sequence models, introduce their applications in asset pricing, and discuss their advantages and limitations. They then perform a comparative analysis of these methods using data on U.S. equities. They demonstrate how sequence modeling benefits investors in general through incorporating complex historical path dependence and that long short-term memory-based models tend to have the best out-of-sample performance.

Cruz Barsce, J., Palombarini, J. A., and Martinez, E. C. (2021). "Automatic tuning of hyper-parameters of reinforcement learning algorithms using Bayesian optimization with behavioral cloning." In: *arXiv e-Print*.
Optimal setting of several hyper-parameters in machine learning algorithms is key to make the most of available data. To this aim, several methods such as evolutionary strategies, random search, Bayesian optimization and heuristic rules of thumb have been proposed. In reinforcement learning (RL), the information content of data gathered by the learning agent while interacting with its environment is heavily dependent on the setting of many hyper-parameters. Therefore, the user of an RL algorithm has to rely on search-based optimization methods, such as grid search or the Nelder-Mead simplex algorithm, that are very inefficient for most RL tasks, slows down significantly the learning curve and leaves to the user the burden of purposefully biasing data gathering. In this work, in order to make an RL algorithm more user-independent, a novel approach for autonomous hyper-parameter setting using Bayesian optimization is proposed. Data from past episodes and different hyper-parameter values are used at a meta-learning level by performing behavioral cloning which helps improving the effectiveness in maximizing a reinforcement learning variant of an acquisition function. Also, by tightly integrating Bayesian optimization in a reinforcement learning agent design, the number of state transitions needed to converge to the optimal policy for a given task is reduced. Computational experiments reveal promising results compared to other manual tweaking and optimization-based approaches which highlights the benefits of changing the algorithm hyper-parameters to increase the information content of generated data.

Cuschieri, N., Vella, V., and Bajada, J. (2021). "TD3-Based Ensemble Reinforcement Learning for Financial Portfolio Optimisation." In: *31st International Conference on Automated Planning and Scheduling*.

Portfolio Selection (PS) is a perennial financial engineering problem that requires determining a strategy for dynamically allocating wealth among a set of portfolio assets to maximise the long-term return. We investigate state-of-the-art Deep Reinforcement Learning (DRL) algorithms that have proven to be ideal for continuous action spaces, mainly Deep Deterministic Policy Gradient (DDPG) and Twin Delayed Deep Deterministic Policy Gradient (TD3), for the PS problem. Furthermore, we investigate the effect of including stock movement prediction indicators in the state representation, and the potential of using an ensemble framework that combines multiple DRL models. Our experiments show that TD3-based models generally perform better than DDPG-based ones when used on real stock trading data. Furthermore, the introduction of additional financial indicators in the state representation was found to have a positive effect over all. Lastly, an ensemble model also showed promising results, consistently beating the baselines used, albeit not all other DRL models.

Cvitanic, J., Kou, S., Wan, X., and Williams, K. L. (2020). "Pi portfolio management: reaching goals while avoiding drawdowns." In: *SSRN e-Print*.
We propose an approach to portfolio selection that explicitly takes into account investors simultaneous investment objectives, such as achieving target return levels and avoiding specific drawdowns. Our approach is consistent with both standard and non-standard risk preferences, such as those of prospect theory. Instead of asking the investor to choose between lotteries, transforming this into an estimate of the risk preferences, and then selecting the portfolio accordingly, we propose to directly offer investors a choice between lotteries" with varying probabilities of experiencing target levels of profits and losses. Our approach enables investors to flexibly assess the effectiveness of portfolio choices under various conditions. We discuss implementation considerations and compare our approach to traditional mean-variance portfolio selection.

D'Eramo, C., Tateo, D., Bonarini, A., Restelli, M., and Peters, J. (2021). "MushroomRL: Simplifying Reinforcement Learning Research." In: *Journal of Machine Learning Research* 22(131), pp. 1–5.
MushroomRL is an open-source Python library developed to simplify the process of implementing and running Reinforcement Learning (RL) experiments. Compared to other available libraries, MushroomRL has been created with the purpose of providing a comprehensive and flexible framework to minimize the effort in implementing and testing novel RL methodologies. The architecture of MushroomRL is built in such a way that every component of a typical RL experiment is already provided, and most of the time users can only focus on the implementation of their own algorithms. MushroomRL is accompanied by a benchmarking suite collecting experimental results of state-of-the-art deep RL algorithms, and allowing to benchmark new ones. The result is a library from which RL researchers can significantly benefit in the critical phase of the empirical analysis of their works. MushroomRL stable code, tutorials, and documentation can be found at `https://github.com/MushroomRL/mushroom-rl`.

Das, S. and Varma, S. (2020). "Dynamic Goals-Based Wealth Management using Reinforcement Learning." In: *Journal of Investment Management* 18 (2), pp. 1–20.
We present a reinforcement learning (RL) algorithm to solve for a dynamically optimal goal-based portfolio. The solution converges to that obtained from dynamic programming. Our approach is model-free and generates a solution that is based on forward simulation, whereas dynamic programming depends on backward recursion. This paper presents a brief overview of the various types of RL. Our example application illustrates how RL may be applied to problems with path-dependency and very large state spaces, which are often encountered in finance.

Das, S. R., Ostrov, D., Radhakrishnan, A., and Srivastav, D. (2020). "Dynamic portfolio allocation in goals-based wealth management." In: *Computational Management Science* 17, pp. 613–640.
We report a dynamic programming algorithm which, given a set of efficient (or even inefficient) portfolios, constructs an optimal portfolio trading strategy that maximizes the probability of attaining an investor's specified target wealth at the end of a designated time horizon. Our algorithm also accommodates periodic infusions or withdrawals of cash with no degradation to the dynamic portfolio's performance or runtime. We explore the sensitivity of the terminal wealth distribution to restricting the segment of the efficient frontier available to the investor. Since our algorithm's optimal strategy can be on the efficient frontier and is driven by an investor's wealth and goals, it soundly beats the performance of target date funds in attaining investors' goals. These optimal goals-based wealth management strategies are useful for independent financial advisors to implement behavioral-based FinTech offerings and for robo-advisors.

Das, S. R., Ostrov, D., Radhakrishnan, A., and Srivastav, D. (2022a). "Dynamic optimization for multi-goals wealth management." In: *Journal of Banking & Finance* 140 (106192).

We develop a dynamic programming methodology that seeks to maximize investor outcomes over multiple, potentially competing goals (such as upgrading a home, paying college tuition, or maintaining an income stream in retirement), even when financial resources are limited. Unlike Monte Carlo approaches currently in wide use in the wealth management industry, our approach uses investor preferences to dynamically make the optimal determination for fulfilling or not fulfilling each goal and for selecting the investor's investment portfolio. This can be computed quickly, even for numerous investor goals spread over different or concurrent time periods, where each goal may be all-or-nothing or may allow for partial fulfillment. The probabilities of attaining each (full or partial) goal under the optimal scenario are also computed, so the investor can ensure the algorithm accurately reflects their preference for the relative importance of each of their goals. This approach vastly outperforms static portfolio strategies and target-date funds, widely used in the wealth management industry.

Das, S. R., Ostrov, D. N., Casanova, A., Radhakrishnan, A., and Srivastav, D. (2021). "Optimal Goals-Based Investment Strategies For Switching Between Bull and Bear Markets." In: *SSRN e-Print*.
We apply dynamic programming to solve a long-horizon fund choice problem, given that the underlying market can switch between different regimes. The objective function is based on reaching a target level of wealth, following the paradigm of goal-based investing. In a world with a good regime (e.g., a bull market) and a bad regime (e.g., a bear market), we find that an investor who is cognizant of regime switching can potentially do much better over time than an investor who assumes there is only one regime. However, there is a caveat—an investor must be able to predict the regime they are in with reasonable levels of confidence, and if not, they are in fact worse off than an investor who assumes just one regime. Using data from recent history, we find that investors may be better off not switching from existing single-regime models to more complex multiple-regime models.

Das, S. R., Ostrov, D. N., Casanova, A., Radhakrishnan, A., and Srivastav, D. (2022b). "Optimal Goals-Based Investment Strategies For Switching Between Bull and Bear Markets." In: *The Journal of Wealth Management* 24(4), pp. 8–36.
We solve a dynamic, long-horizon, goals-based wealth management problem, given different investment regimes. In a world with a good regime (bull market) and a bad regime (bear market), an investor who is cognizant that regime switching occurs has the potential to do better than an investor who assumes only one regime. However, models with more than one regime incur the additional risk of regime uncertainty. Investors must be able to predict which regime is governing the market with reasonable levels of confidence, or they can be worse off than investors who assume just one regime. Using data from recent history, we develop a framework that determines how accurate regime prediction needs to be to achieve gains from a regime-cognizant goals-based investing approach.

Das, S. R. and Ross, G. (2021). "The Role of Options in Goals-Based Wealth Management." In: *SSRN e-Print*.
We develop a facile methodology using dynamic programming for goals-based wealth management over long horizons where rebalancing uses the standard securities and also derivative securities. A kernel density estimation approach is developed to accommodate any number of derivative assets, solving a high dimensional problem with fast computation. The approach easily accommodates skewed and fat-tailed distributions. Portfolio performance is much better with the use of options, especially for investors with aggressive goals.

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.
Any investment process should start with a thorough understanding of the investor problem. Individual investors do not need investment products with alleged superior performance; they need investment solutions that can help them meet their goals subject to prevailing dollar and risk budget constraints. In this new publication, EDHEC-Risk Institute develops a general operational framework that can be used by financial advisors to allow individual investors to optimally allocate to categories of risks they face across all life stages and wealth segments so as to achieve personally meaningful financial goals.

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.
This article is concerned with the use of currently available technology to offer individuals, financial advisors, and pension fund financial planners detailed prospective financial plans tailored to an individual's financial goals and obligations. By taking account of all an individual's prospective cash flows, including servicing current liabilities, and simultaneously optimizing prospective spending, saving, asset allocation, tax, and insurance, etc. using dynamic stochastic optimization, the authors compare the results of their goal-based

fully dynamic strategy with the financial advisory industry's representative current best practices. These include piecemeal fixed-allocation portfolios for specific goals, target-date retirement funds, and fixed real-income post-retirement financial products, all using Markowitz mean-variance optimization based on the very general goal of minimizing portfolio volatility for a specific portfolio expected return over a finite horizon. Making use of the same data and marketcalibrated Monte Carlo stochastic simulation for all the alternative portfolio strategies, the authors find that flexibility is of key importance for both individual portfolio and spending decisions. The authors measure superiority by the certainty-equivalent increase in expected utility of individual lifetime consumption (gamma) and the extra initial capital required by an individual to put the dominated strategy on the same expected-utility footing as the optimal dynamic strategy (initial capital gap). They find that the adaptive dynamic goal-based portfolio strategy's performance is far superior to all the industry's Markowitz-based approaches. These empirical results should put paid to the commonly held view that the extra complexity of holistic dynamic stochastic models is not worth the marginal extra value obtained from their use.

Denault, M. and Simonato, J.-G. (2022). "A note on a dynamic goal-based wealth management problem." In: *Finance Research Letters* 46 (Part B) (102404).
This short note suggests two improvements for solving the goal-based wealth management problem proposed by Das, Ostrov, Radhakrishnan and Srivastav (2020). The first suggestion smoothes and improves the convergence of the approximate solutions towards the underlying, continuous solution either by using analytic solutions at the penultimate time point or by adjusting the wealth grid. The second suggestion pertains to fast matrix products and is purely computational but has a large impact on the time required to solve the problem. We also propose a more consistent approximation for the calculation of the return parameters.

Diris, B., Palm, F., and Schotman, P. (2015). "Long-Term Strategic Asset Allocation: An Out-of-Sample Evaluation." In: *Management Science* 61(9), pp. 2185–2202.
We evaluate the out-of-sample performance of a long-term investor who follows an optimized dynamic trading strategy. Although the dynamic strategy is able to benefit from predictability out-of-sample, a short-term investor using a single-period market timing strategy would have realized an almost identical performance. The value of intertemporal hedge demands in strategic asset allocation appears negligible. The result is caused by the estimation error in predicting the predictors. A myopic investor only needs to predict one-period-ahead expected returns, but hedge demands also require accurate predictions of the predictor variables. To reduce the problem of errors in optimized portfolio weights, we consider Bayesian procedures. Myopic and dynamic portfolios are similarly affected by such modifications, and differences in performance become even smaller.

Dixon, M. F., Halperin, I., and Bilokon, P. (2020). *Machine Learning in Finance: from theory to practice*. Springer International Publishing. 548 pp.
This book introduces machine learning methods in finance. It presents a unified treatment of machine learning and various statistical and computational disciplines in quantitative finance, such as financial econometrics and discrete time stochastic control, with an emphasis on how theory and hypothesis tests inform the choice of algorithm for financial data modeling and decision making. With the trend towards increasing computational resources and larger datasets, machine learning has grown into an important skillset for the finance industry. This book is written for advanced graduate students and academics in financial econometrics, mathematical finance and applied statistics, in addition to quants and data scientists in the field of quantitative finance. Machine Learning in Finance: From Theory to Practice is divided into three parts, each part covering theory and applications. The first presents supervised learning for cross-sectional data from both a Bayesian and frequentist perspective. The more advanced material places a firm emphasis on neural networks, including deep learning, as well as Gaussian processes, with examples in investment management and derivative modeling. The second part presents supervised learning for time series data, arguably the most common data type used in finance with examples in trading, stochastic volatility and fixed income modeling. Finally, the third part presents reinforcement learning and its applications in trading, investment and wealth management. Python code examples are provided to support the readers' understanding of the methodologies and applications. The book also includes more than 80 mathematical and programming exercises, with worked solutions available to instructors. As a bridge to research in this emergent field, the final chapter presents the frontiers of machine learning in finance from a researcher's perspective, highlighting how many well-known concepts in statistical physics are likely to emerge as important methodologies for machine learning in finance.

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

We present a reinforcement learning approach to goal based wealth management problems such as optimization of retirement plans or target dated funds. In such problems, an investor seeks to achieve a financial goal by making periodic investments in the portfolio while being employed, and periodically draws from the account when in retirement, in addition to the ability to re-balance the portfolio by selling and buying different assets (e.g. stocks). Instead of relying on a utility of consumption, we present G-Learner: a reinforcement learning algorithm that operates with explicitly defined one-step rewards, does not assume a data generation process, and is suitable for noisy data. Our approach is based on G-learning (Fox et al., 2015) - a probabilistic extension of the Q-learning method of reinforcement learning. In this paper, we demonstrate how G-learning, when applied to a quadratic reward and Gaussian reference policy, gives an entropy-regulated Linear Quadratic Regulator (LQR). This critical insight provides a novel and computationally tractable tool for wealth management tasks which scales to high dimensional portfolios. In addition to the solution of the direct problem of G-learning, we also present a new algorithm, GIRL, that extends our goal-based G-learning approach to the setting of Inverse Reinforcement Learning (IRL) where rewards collected by the agent are not observed, and should instead be inferred. We demonstrate that GIRL can successfully learn the reward parameters of a G-Learner agent and thus imitate its behavior. Finally, we discuss potential applications of the G-Learner and GIRL algorithms for wealth management and robo-advising.

Dixon, M. F. and Halperin, I. (2021). "Goal-based wealth management with generative reinforcement learning." In: *Risk (Cutting edge)*.
A combination of machine learning techniques provides multi-period portfolio optimisation. Matthew Dixon and Igor Halperin develop a reinforcement learning (RL) approach to goal-based wealth management problems such as optimisation of retirement plans or target date funds. They present G-Learner: a reinforcement learning algorithm that does not assume a data generation process and is suitable for noisy data. Their approach is based on G-learning, a probabilistic extension of the Q-learning method of reinforcement learning. In addition to G-Learners, which solves the direct RL problem, they develop GIRL, a G-learning inverse RL algorithm to infer the investor reward function from the observed trading actions.

Dong, L., Gao, G., Li, Y., and Wen, Y. (2021a). "Baconian: A Unified Open source Framework for Model-Based Reinforcement Learning." In: *arXiv e-Print*.
Model-Based Reinforcement Learning (MBRL) is one category of Reinforcement Learning (RL) methods which can improve sampling efficiency by modeling and approximating system dynamics. It has been widely adopted in the research of robotics, autonomous driving, etc. Despite its popularity, there still lacks some sophisticated and reusable opensource frameworks to facilitate MBRL research and experiments. To fill this gap, we develop a flexible and modularized framework, Baconian, which allows researchers to easily implement a MBRL testbed by customizing or building upon our provided modules and algorithms. Our framework can free the users from re-implementing popular MBRL algorithms from scratch thus greatly saves the users' efforts.

Dong, Z., Huang, S., Ma, S., and Qian, Y. (2021b). "Factor Representation and Decision Making in Stock Markets Using Deep Reinforcement Learning." In: *arXiv e-Print*.
Deep Reinforcement learning is a branch of unsupervised learning in which an agent learns to act based on environment state in order to maximize its total reward. Deep reinforcement learning provides good opportunity to model the complexity of portfolio choice in high-dimensional and data-driven environment by leveraging the powerful representation of deep neural networks. In this paper, we build a portfolio management system using direct deep reinforcement learning to make optimal portfolio choice periodically among S&P500 underlying stocks by learning a good factor representation (as input). The result shows that an effective learning of market conditions and optimal portfolio allocations can significantly outperform the average market.

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.
The authors propose models for the solution of the fundamental problem of option replication subject to discrete trading, round lotting, and nonlinear transaction costs using state-of-the-art methods in deep reinforcement learning (DRL), including deep Q-learning, deep Q-learning with Pop-Art, and proximal policy optimization (PPO). Each DRL model is trained to hedge a whole range of strikes, and no retraining is needed when the user changes to another strike within the range. The models are general, allowing the user to plug in any option pricing and simulation library and then train them with no further modifications to hedge arbitrary option portfolios. Through a series of simulations, the authors show that the DRL models

learn similar or better strategies as compared to delta hedging. Out of all models, PPO performs the best in terms of profit and loss, training time, and amount of data needed for training.

Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Gowal, S., and Hester, T. (2021). "An empirical investigation of the challenges of real-world reinforcement learning." In: *arXiv e-Print*.
Reinforcement learning (RL) has proven its worth in a series of artificial domains, and is beginning to show some successes in real-world scenarios. However, much of the research advances in RL are hard to leverage in real-world systems due to a series of assumptions that are rarely satisfied in practice. In this work, we identify and formalize a series of independent challenges that embody the difficulties that must be addressed for RL to be commonly deployed in real-world systems. For each challenge, we define it formally in the context of a Markov Decision Process, analyze the effects of the challenge on state-of-the-art learning algorithms, and present some existing attempts at tackling it. We believe that an approach that addresses our set of proposed challenges would be readily deployable in a large number of real world problems. Our proposed challenges are implemented in a suite of continuous control environments called realworldrl-suite which we propose an as an open-source benchmark.

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.
Selection bias under multiple testing is a serious problem. From a practitioner perspective, failure to disclose the impact of multiple tests of a proposed investment strategy to clients and senior management can lead to the adoption of a false discovery. Clients will lose money, senior management will misallocate resources, and the firm may be exposed to reputational, legal, and regulatory risks. From the perspective of academic journals that publish evidence supporting an investment strategy, the failure to address selection bias under multiple testing threatens to invalidate large portions of the literature in empirical finance. In this article, the authors propose a template that practitioners should use to fairly disclose multiple tests involved in an alleged discovery when pitching strategies to clients and senior management. The same template could be used by contributors to academic journals so that referees, and ultimately readers, can assess the strategy. By disclosing this information, those who are charged with making the final decision about a discovery can evaluate the probability that the purported discovery is false.

Fedus, W., Gelada, C., Bengio, Y., Bellemare, M. G., and Larochelle, H. (2019). "Hyperbolic Discounting and Learning over Multiple Horizons." In: *arXiv e-Print*.
Reinforcement learning (RL) typically defines a discount factor as part of the Markov Decision Process. The discount factor values future rewards by an exponential scheme that leads to theoretical convergence guarantees of the Bellman equation. However, evidence from psychology, economics and neuroscience suggests that humans and animals instead have hyperbolic time-preferences. In this work we revisit the fundamentals of discounting in RL and bridge this disconnect by implementing an RL agent that acts via hyperbolic discounting. We demonstrate that a simple approach approximates hyperbolic discount functions while still using familiar temporal-difference learning techniques in RL. Additionally, and independent of hyperbolic discounting, we make a surprising discovery that simultaneously learning value functions over multiple time-horizons is an effective auxiliary task which often improves over a strong value-based RL agent, Rainbow.

Fischer, T. G. (2018). *Reinforcement learning in financial markets - a survey*. Tech. rep. Friedrich-Alexander University.
The advent of reinforcement learning (RL) in financial markets is driven by several advantages inherent to this field of artificial intelligence. In particular, RL allows to combine the "prediction" and the "portfolio construction" task in one integrated step, thereby closely aligning the machine learning problem with the objectives of the investor. At the same time, important constraints, such as transaction costs, market liquidity, and the investor's degree of risk-aversion, can be conveniently taken into account. Over the past two decades, and albeit most attention still being devoted to supervised learning methods, the RL research community has made considerable advances in the finance domain. The present paper draws insights from almost 50 publications, and categorizes them into three main approaches, i.e., critic-only approach, actor-only approach, and actor-critic approach. Within each of these categories, the respective contributions are summarized and reviewed along the representation of the state, the applied reward function, and the action space of the agent. This cross-sectional perspective allows us to identify recurring design decisions as well as potential levers to improve the agent's performance. Finally, the individual strengths and weaknesses of each approach are discussed, and directions for future research are pointed out.

Fleiss, A., Kumaar, A., Rida, A., Shin, J., Lai, X., Fant, V., Chen, J., and Li, A. (2021). "Deep Reinforcement Learning and Feature Extraction For Constructing Alpha Generating Equity Portfolios." In: *SSRN e-Print*.
The ambition of this paper is to catch hidden information inside the Securities and Exchange Commission's (SEC)13F public holding data in order to construct an equity portfolio that maximizes returns. The 13F ling data give us the quarterly stock trading decisions of included funds, but we're not given any insight on how they made their decisions or if information has been shared between funds. To remedy this lack of knowledge, this paper used feature extraction in order to lter out the best performing funds through several criteria. We propose a method employing powerful machine learning techniques (Deep Reinforcement Learning) to try to catch the missing pieces of information behind the decision process and use them as a prediction tool to construct quarterly equity portfolios. This approach reached an annualized return of 21% with a sharpe ratio of 1.8 outperforming the S&P 500 both in returns and stability through historical backtesting.

Forsyth, P. A., Vetzal, K. R., and Westmacott, G. (2021). "Optimal control of the decumulation of a retirement portfolio with variable spending and dynamic asset allocation." In: *arXiv e-Print*.
We extend the Annually Recalculated Virtual Annuity (ARVA) spending rule for retirement savings decumulation to include a cap and a floor on withdrawals. With a minimum withdrawal constraint, the ARVA strategy runs the risk of depleting the investment portfolio. We determine the dynamic asset allocation strategy which maximizes a weighted combination of expected total withdrawals (EW) and expected shortfall (ES), defined as the average of the worst five per cent of the outcomes of real terminal wealth. We compare the performance of our dynamic strategy to simpler alternatives which maintain constant asset allocation weights over time accompanied by either our same modified ARVA spending rule or withdrawals that are constant over time in real terms. Tests are carried out using both a parametric model of historical asset returns as well as bootstrap resampling of historical data. Consistent with previous literature that has used different measures of reward and risk than EW and ES, we find that allowing some variability in withdrawals leads to large improvements in efficiency. However, unlike the prior literature, we also demonstrate that further significant enhancements are possible through incorporating a dynamic asset allocation strategy rather than simply keeping asset allocation weights constant throughout retirement.

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.
Deep reinforcement learning is the combination of reinforcement learning (RL) and deep learning. This field of research has been able to solve a wide range of complex decisionmaking tasks that were previously out of reach for a machine. Thus, deep RL opens up many new applications in domains such as healthcare, robotics, smart grids, finance, and many more. This manuscript provides an introduction to deep reinforcement learning models, algorithms and techniques. Particular focus is on the aspects related to generalization and how deep RL can be used for practical applications. We assume the reader is familiar with basic machine learning concepts.

Freymuth, N., Becker, P., and Neumann, G. (2021). "Versatile Inverse Reinforcement Learning via Cumulative Rewards." In: *arXiv e-Print*.
Inverse Reinforcement Learning infers a reward function from expert demonstrations, aiming to encode the behavior and intentions of the expert. Current approaches usually do this with generative and uni-modal models, meaning that they encode a single behavior. In the common setting, where there are various solutions to a problem and the experts show versatile behavior this severely limits the generalization capabilities of these methods. We propose a novel method for Inverse Reinforcement Learning that overcomes these problems by formulating the recovered reward as a sum of iteratively trained discriminators. We show on simulated tasks that our approach is able to recover general, high-quality reward functions and produces policies of the same quality as behavioral cloning approaches designed for versatile behavior.

Fu, J., Tacchetti, A., Perolat, J., and Bachrach, Y. (2021). "Evaluating Strategic Structures in Multi-Agent Inverse Reinforcement Learning." In: *Journal of Artificial Intelligence Research* 71, pp. 925–951.
A core question in multi-agent systems is understanding the motivations for an agent's actions based on their behavior. Inverse reinforcement learning provides a framework for extracting utility functions from observed agent behavior, casting the problem as finding domain parameters which induce such a behavior from rational decision makers. We show how to efficiently and scalably extend inverse reinforcement learning to multi-agent settings, by reducing the multi-agent problem to N single-agent problems while still satisfying rationality conditions such as strong rationality. However, we observe that rewards learned naively tend to lack insightful structure, which causes them to produce undesirable behavior when optimized in games

with different players from those encountered during training. We further investigate conditions under which rewards or utility functions can be precisely identified, on problem domains such as normal-form and Markov games, as well as auctions, where we show we can learn reward functions that properly generalize to new settings.

Fujita, Y., Kataoka, T., Nagarajan, P., and Ishikawa, T. (2021). "ChainerRL: A Deep Reinforcement Learning Library." In: *Journal of Machine Learning Research* 22(77), pp. 1–14.
In this paper, we introduce ChainerRL, an open-source deep reinforcement learning (DRL) library built using Python and the Chainer deep learning framework. ChainerRL implements a comprehensive set of DRL algorithms and techniques drawn from state-of-the-art research in the field. To foster reproducible research, and for instructional purposes, ChainerRL provides scripts that closely replicate the original papers' experimental settings and reproduce published benchmark results for several algorithms. Lastly, ChainerRL offers a visualization tool that enables the qualitative inspection of trained agents. The ChainerRL source code can be found on GitHub: `https://github.com/chainer/chainerrl`.

Garau-Luis, J. J., Crawley, E., and Cameron, B. (2021). "Evaluating the progress of Deep Reinforcement Learning in the real world: aligning domain-agnostic and domain-specific research." In: *arXiv e-Print*.
Deep Reinforcement Learning (DRL) is considered a potential framework to improve many real-world autonomous systems; it has attracted the attention of multiple and diverse fields. Nevertheless, the successful deployment in the real world is a test most of DRL models still need to pass. In this work we focus on this issue by reviewing and evaluating the research efforts from both domain-agnostic and domain-specific communities. On one hand, we offer a comprehensive summary of DRL challenges and summarize the different proposals to mitigate them; this helps identifying five gaps of domain-agnostic research. On the other hand, from the domain-specific perspective, we discuss different success stories and argue why other models might fail to be deployed. Finally, we take up on ways to move forward accounting for both perspectives.

Gasperov, B., Saric, F., Begusic, S., and Kostanjcar, Z. (2020). "Adaptive rolling window selection for minimum variance portfolio estimation based on reinforcement learning." In: *43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*. IEEE.
When allocating wealth to a set of financial assets, portfolio optimization techniques are used to select optimal portfolio allocations for given investment goals. Among benchmark portfolios commonly used in modern portfolio theory, the global minimum variance portfolio is becoming increasingly popular with investors due to its relatively good performance which stems from both the low-volatility anomaly and the avoidance of the estimation of first moments i.e. mean returns. However, estimates of minimum variance portfolio weights significantly depend on the size of the rolling window used for estimation, especially considering the non-stationarity of the underlying market dynamics. In this paper, we use a model-free policy-based reinforcement learning framework in order to directly and adaptively determine the optimal size of the rolling window. Training is done on a subset of trading stocks from the NYSE. The resulting agent achieves superior performance when compared against multiple benchmarks, including those with fixed rolling window sizes.

Gašperov, B., Begušić, S., Šimović, P. P., and Kostanjčar, Z. (2021). "Reinforcement Learning Approaches to Optimal Market Making." In: *Mathematics* 9(21), p. 2689.
Market making is the process whereby a market participant, called a market maker, simultaneously and repeatedly posts limit orders on both sides of the limit order book of a security in order to both provide liquidity and generate profit. Optimal market making entails dynamic adjustment of bid and ask prices in response to the market maker's current inventory level and market conditions with the goal of maximizing a risk-adjusted return measure. This problem is naturally framed as a Markov decision process, a discrete-time stochastic (inventory) control process. Reinforcement learning, a class of techniques based on learning from observations and used for solving Markov decision processes, lends itself particularly well to it. Recent years have seen a very strong uptick in the popularity of such techniques in the field, fueled in part by a series of successes of deep reinforcement learning in other domains. The primary goal of this paper is to provide a comprehensive and up-to-date overview of the current state-of-the-art applications of (deep) reinforcement learning focused on optimal market making. The analysis indicated that reinforcement learning techniques provide superior performance in terms of the risk-adjusted return over more standard market making strategies, typically derived from analytical models.

Glanois, C., Weng, P., Zimmer, M., Li, D., Yang, T., Hao, J., and Liu, W. (2022). "A Survey on Interpretable Reinforcement Learning." In: *arXiv e-Print*.

Although deep reinforcement learning has become a promising machine learning approach for sequential decision-making problems, it is still not mature enough for high-stake domains such as autonomous driving or medical applications. In such contexts, a learned policy needs for instance to be interpretable, so that it can be inspected before any deployment (e.g., for safety and verifiability reasons). This survey provides an overview of various approaches to achieve higher interpretability in reinforcement learning (RL). To that aim, we distinguish interpretability (as a property of a model) and explainability (as a post-hoc operation, with the intervention of a proxy) and discuss them in the context of RL with an emphasis on the former notion. In particular, we argue that interpretable RL may embrace different facets: interpretable inputs, interpretable (transition/reward) models, and interpretable decision-making. Based on this scheme, we summarize and analyze recent work related to interpretable RL with an emphasis on papers published in the past 10 years. We also discuss briefly some related research areas and point to some potential promising research directions.

Greiner, S. P. and Stoyanov, S. V. (2019). "Portfolio scoring by expected risk premium." In: *The Journal of Portfolio Management* 45(4), pp. 83–90.
In this article, the authors discuss a general method for ranking portfolios that places few limitations on the portfolio constituents other than using publicly traded assets. The ranking scores reflect the expected reward investors would require for accepting the risks of the portfolio in the context of an asset pricing framework. The scores are computed through a factor model that acknowledges the factor return correlations. The authors illustrate the approach with a large universe of exchange-traded funds assuming a linear model with Fama-French-Carhart factors wherein factor premiums (i.e., expected returns) are proportional to factor volatilities. The empirical analysis implies that the most significant factors from the Fama-French-Carhart factor set driving the premiums are the market and the momentum factors.

Gu, S. (2021). "Deep Reinforcement Learning with Function Properties in Mean Reversion Strategies." In: *arXiv e-Print*.
With the recent advancement in Deep Reinforcement Learning in the gaming industry, we are curious if the same technology would work as well for common quantitative financial problems. In this paper, we will investigate if an off-the-shelf library developed by OpenAI can be easily adapted to mean reversion strategy. Moreover, we will design and test to see if we can get better performance by narrowing the function space that the agent needs to search for. We achieve this through augmenting the reward function by a carefully picked penalty term.

Guan, M. and Liu, X.-Y. (2021). "Explainable Deep Reinforcement Learning for Portfolio Management: An Empirical Approach." In: *arXiv e-Print*.
Deep reinforcement learning (DRL) has been widely studied in the portfolio management task. However, it is challenging to understand a DRL-based trading strategy because of the black-box nature of deep neural networks. In this paper, we propose an empirical approach to explain the strategies of DRL agents for the portfolio management task. First, we use a linear model in hindsight as the reference model, which finds the best portfolio weights by assuming knowing actual stock returns in foresight. In particular, we use the coefficients of a linear model in hindsight as the reference feature weights. Secondly, for DRL agents, we use integrated gradients to define the feature weights, which are the coefficients between reward and features under a linear regression model. Thirdly, we study the prediction power in two cases, single-step prediction and multi-step prediction. In particular, we quantify the prediction power by calculating the linear correlations between the feature weights of a DRL agent and the reference feature weights, and similarly for machine learning methods. Finally, we evaluate a portfolio management task on Dow Jones 30 constituent stocks during 01/01/2009 to 09/01/2021. Our approach empirically reveals that a DRL agent exhibits a stronger multi-step prediction power than machine learning methods.

Guidolin, M., Hansen, E., and Lozano-Banda, M. (2018). "Portfolio performance of linear SDF models: an out-of-sample assessment." In: *Quantitative Finance* 18(8), pp. 1425–1436.
We evaluate linear stochastic discount factor models using an ex-post portfolio metric: the realized out-of-sample Sharpe ratio of mean-variance portfolios backed by alternative linear factor models. Using a sample of monthly US portfolio returns spanning the period 1968-2016, we find evidence that multifactor linear models have better empirical properties than the CAPM, not only when the cross-section of expected returns is evaluated in-sample, but also when they are used to inform one-month ahead portfolio selection. When we compare portfolios associated to multifactor models with mean-variance decisions implied by the single-factor CAPM, we document statistically significant differences in Sharpe ratios of up to 10 percent. Linear multifactor models that provide the best in-sample fit also yield the highest realized Sharpe ratios.

Guo, D. (2019). "A Statistical Response to Challenges in Vast Portfolio Selection." PhD thesis. University of Waterloo.
The thesis is written in response to emerging issues brought about by an increasing number of assets allocated in a portfolio and seeks answers to puzzling empirical findings in the portfolio management area. Over the years, researchers and practitioners working in the portfolio optimization area have been concerned with estimation errors in the first two moments of asset returns. The thesis comprises several related chapters on our statistical inquiry into this subject. Chapter 1 of the thesis contains an introduction to what will be reported in the remaining chapters. A few well-known covariance matrix estimation methods in the literature involve adjustment of sample eigenvalues. Chapter 2 of the thesis examines the effects of sample eigenvalue adjustment on the out-of-sample performance of a portfolio constructed from the sample covariance matrix.

Guo, D., Boyle, P. P., Weng, C., and Wirjanto, T. S. (2019). "When Does The 1/N Rule Work?" In: *SSRN e-Print*.
We propose a "1/N favorability index" to measure how favorable a market is to holding a 1/N portfolio. This index reflects the extent of difficulty for an optimized portfolio to outperform the 1/N portfolio in a specific market. A single-factor model predicts that bull markets are accompanied by a high 1/N favorability index and vice versa. We validate the model implication that the 1/N portfolio is more difficult to beat in bull markets using stock return datasets from a number of countries as well as the classic datasets used by DeMiguel et al. (2009). Our results imply that the reported good performance of the 1/N portfolio in the US equity market can be partially attributed to the long-run bullish trend in the market which gives rise to the high favorability of the market to the 1/N portfolio.

Guo, I., Langrené, N., Loeper, G., and Ning, W. (2022). "Portfolio optimization with a prescribed terminal wealth distribution." In: *Quantitative Finance*, pp. 1–15.
This paper studies a portfolio allocation problem, where the goal is to reach a prescribed wealth distribution at a final time. We study this problem with the tools of optimal mass transport. We provide a dual formulation which is solved with a gradient descent algorithm. This involves solving an associated Hamilton-Jacobi-Bellman and Fokker-Planck equations with a finite difference method. Numerical examples for various prescribed terminal distributions are given, showing that we can successfully reach attainable targets. We then consider adding consumption during the investment process, to take into account distributions that are either not attainable, or sub-optimal.

Haley, M. R. (2017). "K-fold cross validation performance comparisons of six naive portfolio selection rules: how naive can you be and still have successful out-of-sample portfolio performance?" In: *Annals of Finance* 13, pp. 341–353.
Recent research reports that optimal portfolio selection models often perform worse than equal-weight naive diversification in out-of-sample testing. This paper extends this line of inquiry by comparing the out-of-sample performance of the equal-weight naive strategy to the out-of-sample performance of five alternative naive strategies, each of which derives from a simple heuristic that does not require any optimization. Out-of-sample portfolio performance is assessed by mean, standard deviation, skewness, and Sharpe ratio; k-fold cross validation is used as the out-of-sample testing mechanism. The results indicate that the proposed naive heuristic rules exhibit strong out-of-sample performance, in most cases superior to the equal-weight naive strategy. These findings are consequential for at least two reasons: first, if these simple heuristic-based rules outperform the equal-weight naive strategy, then by transitivity they can outperform the mean-variance- and shortfall-optimal portfolio rules that have been shown in the literature to be inferior to the equal-weight naive rule, which further emphasizes the out-of-sample fragility of "optimal"methods; and second, among naive diversification strategies, some appear more robust in out-of-sample testing than others, hence the proposed methods may be useful when forming mixed portfolio selection models wherein a naive strategy is combined with an optimal strategy to improve performance.

Halperin, I. (2019). "The QLBS Q-Learner goes NuQLear: fitted Q iteration, inverse RL, and option portfolios." In: *Quantitative Finance* 19(9), pp. 1543–1553.
The QLBS model is a discrete-time option hedging and pricing model that is based on Dynamic Programming (DP) and Reinforcement Learning (RL). It combines the famous Q-Learning method for RL with the Black-Scholes (-Merton) (BSM) model's idea of reducing the problem of option pricing and hedging to the problem of optimal rebalancing of a dynamic replicating portfolio for the option, which is made of a stock and cash. Here we expand on several NuQLear (Numerical Q-Learning) topics with the QLBS model. First, we investigate the performance of Fitted Q Iteration for an RL (data-driven) solution to the model, and benchmark it versus

a DP (model-based) solution, as well as versus the BSM model. Second, we develop an Inverse Reinforcement Learning (IRL) setting for the model, where we only observe prices and actions (re-hedges) taken by a trader, but not rewards. Third, we outline how the QLBS model can be used for pricing portfolios of options, rather than a single option in isolation, thus providing its own, data-driven and model-independent solution to the (in)famous volatility smile problem of the Black-Scholes model.

Halperin, I. and Feldshteyn, I. (2018). "Market Self-Learning of Signals, Impact and Optimal Trading: Invisible Hand Inference with Free Energy (Or, How We Learned to Stop Worrying and Love Bounded Rationality)." In: *SSRN e-Print*.

We present a simple model of a non-equilibrium self-organizing market where asset prices are partially driven by investment decisions of a bounded-rational agent. The agent acts in a stochastic market environment driven by various exogenous "alpha" signals, agent's own actions (via market impact), and noise. Unlike traditional agent-based models, our agent aggregates all traders in the market, rather than being a representative agent. Therefore, it can be identified with a bounded-rational component of the market itself, providing a particular implementation of an Invisible Hand market mechanism. In such setting, market dynamics are modeled as a fictitious self-play of such bounded-rational market-agent in its adversarial stochastic environment. As rewards obtained by such self-playing market agent are not observed from market data, we formulate and solve a simple model of such market dynamics based on a neuroscience-inspired Bounded Rational Information Theoretic Inverse Reinforcement Learning (BRIT-IRL). This results in effective asset price dynamics with a non-linear mean reversion - which in our model is generated dynamically, rather than being postulated. We argue that our model can be used in a similar way to the Black-Litterman model. In particular, it represents, in a simple modeling framework, market views of common predictive signals, market impacts and implied optimal dynamic portfolio allocations, and can be used to assess values of private signals. Moreover, it allows one to quantify a "market-implied" optimal investment strategy, along with a measure of market rationality. Our approach is numerically light, and can be implemented using standard off-the-shelf software such as TensorFlow.

Halperin, I., Liu, J., and Zhang, X. (2022). "Combining Reinforcement Learning and Inverse Reinforcement Learning for Asset Allocation Recommendations." In: *arXiv e-Print*.

We suggest a simple practical method to combine the human and artificial intelligence to both learn best investment practices of fund managers, and provide recommendations to improve them. Our approach is based on a combination of Inverse Reinforcement Learning (IRL) and RL. First, the IRL component learns the intent of fund managers as suggested by their trading history, and recovers their implied reward function. At the second step, this reward function is used by a direct RL algorithm to optimize asset allocation decisions. We show that our method is able to improve over the performance of individual fund managers.

Hamadanian, P., Schwarzkopf, M., Sen, S., and Alizadeh, M. (2022). "Reinforcement Learning in Time-Varying Systems: an Empirical Study." In: *arXiv e-Print*.

Recent research has turned to Reinforcement Learning (RL) to solve challenging decision problems, as an alternative to hand-tuned heuristics. RL can learn good policies without the need for modeling the environment's dynamics. Despite this promise, RL remains an impractical solution for many real-world systems problems. A particularly challenging case occurs when the environment changes over time, i.e. it exhibits non-stationarity. In this work, we characterize the challenges introduced by non-stationarity and develop a framework for addressing them to train RL agents in live systems. Such agents must explore and learn new environments, without hurting the system's performance, and remember them over time. To this end, our framework (1) identifies different environments encountered by the live system, (2) explores and trains a separate expert policy for each environment, and (3) employs safeguards to protect the system's performance. We apply our framework to two systems problems: straggler mitigation and adaptive video streaming, and evaluate it against a variety of alternative approaches using real-world and synthetic data. We show that each component of our framework is necessary to cope with non-stationarity.

Hambly, B. M., Xu, R., and Yang, H. (2021). "Recent Advances in Reinforcement Learning in Finance." In: *SSRN e-Print*.

The rapid changes in the finance industry due to the increasing amount of data has revolutionized the techniques on data processing and data analysis and brought new theoretical and computational challenges. In contrast to classical stochastic control theory and other analytical approaches for solving financial decision-making problems that heavily reply on model assumptions, new developments from reinforcement learning (RL) are able to make full use of the large amount of financial data with fewer model assumptions and

to improve decisions in complex financial en- environments. This survey paper aims to review the recent developments and use of RL approaches in finance. We give an introduction to Markov decision processes, which is the setting for many of the commonly used RL approaches. Various algorithms are then introduced with a focus on value and policy based methods that do not require any model assumptions. Connections are made with neural networks to extend the framework to encompass deep RL algorithms. Our survey concludes by discussing the application of these RL algorithms in a variety of decision-making problems in finance, including optimal execution, portfolio optimization, option pricing and hedging, market making, smart order routing, and robo-advising.

Harvey, C. R., Liu, Y., and Saretto, A. (2020). "An Evaluation of Alternative Multiple Testing Methods for Finance Applications." In: *The Review of Asset Pricing Studies* 10(2), pp. 199–248.
In almost every area of empirical finance, researchers confront multiple tests. One high-profile example is the identification of outperforming investment managers, many of whom beat their benchmarks purely by luck. Multiple testing methods are designed to control for luck. Factor selection is another glaring case in which multiple tests are performed, but numerous other applications do not receive as much attention. One important example is a simple regression model testing five variables. In this case, because five variables are tried, a t-statistic of 2.0 is not enough to establish significance. Our paper provides a guide to various multiple testing methods and details a number of applications. We provide simulation evidence on the relative performance of different methods across a variety of testing environments. The goal of our paper is to provide a menu that researchers can choose from to improve inference in financial economics.

Hayes, C. F., Radulescu, R., Bargiacchi, E., Kallstrom, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L. M., Dazeley, R., Heintz, F., Howley, E., Irissappane, A. A., Mannion, P., Nowe, A., Ramos, G., Restelli, M., Vamplew, P., and Roijers, D. M. (2021). "A Practical Guide to Multi-Objective Reinforcement Learning and Planning." In: *arXiv e-Print*.
Real-world decision-making tasks are generally complex, requiring trade-offs between multiple, often conflicting, objectives. Despite this, the majority of research in reinforcement learning and decision-theoretic planning either assumes only a single objective, or that multiple objectives can be adequately handled via a simple linear combination. Such approaches may oversimplify the underlying problem and hence produce suboptimal results. This paper serves as a guide to the application of multi-objective methods to difficult problems, and is aimed at researchers who are already familiar with single-objective reinforcement learning and planning methods who wish to adopt a multi-objective perspective on their research, as well as practitioners who encounter multi-objective decision problems in practice. It identifies the factors that may influence the nature of the desired solution, and illustrates by example how these influence the design of multi-objective decision-making systems for complex problems.

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.
Two tests can help asset managers to develop more robust investment strategies: an impact test and a survival test. Both tests complement the backtest, in which one checks how a proposed investment strategy would have performed in the past. The impact test considers the performance of the strategy when assets under management grow (crowdedness), and it checks the impact that growth in assets under management in competing strategies has on the proposed strategy (cross impact). The survival test considers the effect of the long-term evolution of assets under management in competition for market capital. Using Shiller S&P 500 index and bond market data, we show that time-series momentum (relative strength) performs best in the backtest and the impact test but that an expected relative cash-flow rule (relative dividend yield) has the best long-term survival properties.

Heuillet, A., Couthouis, F., and Díaz-Rodríguez, N. (2021). "Explainability in deep reinforcement learning." In: *Knowledge-Based Systems* 214, p. 106685.
A large set of the explainable Artificial Intelligence (XAI) literature is emerging on feature relevance techniques to explain a deep neural network (DNN) output or explaining models that ingest image source data. However, assessing how XAI techniques can help understand models beyond classification tasks, e.g. for reinforcement learning (RL), has not been extensively studied. We review recent works in the direction to attain Explainable Reinforcement Learning (XRL), a relatively new subfield of Explainable Artificial Intelligence, intended to be used in general public applications, with diverse audiences, requiring ethical, responsible and trustable algorithms. In critical situations where it is essential to justify and explain the agent's behaviour, better explainability and interpretability of RL models could help gain scientific insight on the inner workings

of what is still considered a black box. We evaluate mainly studies directly linking explainability to RL, and split these into two categories according to the way the explanations are generated: transparent algorithms and post-hoc explainability. We also review the most prominent XAI works from the lenses of how they could potentially enlighten the further deployment of the latest advances in RL, in the demanding present and future of everyday problems.

Hieu, L. T. (2020). "Deep Reinforcement Learning for Stock Portfolio Optimization." In: *arXiv e-Print*.
Stock portfolio optimization is the process of constant re-distribution of money to a pool of various stocks. In this paper, we will formulate the problem such that we can apply Reinforcement Learning for the task properly. To maintain a realistic assumption about the market, we will incorporate transaction cost and risk factor into the state as well. On top of that, we will apply various state-of-the-art Deep Reinforcement Learning algorithms for comparison. Since the action space is continuous, the realistic formulation were tested under a family of state-of-the-art continuous policy gradients algorithms: Deep Deterministic Policy Gradient (DDPG), Generalized Deterministic Policy Gradient (GDPG) and Proximal Policy Optimization (PPO), where the former two perform much better than the last one. Next, we will present the end-to-end solution for the task with Minimum Variance Portfolio Theory for stock subset selection, and Wavelet Transform for extracting multi-frequency data pattern. Observations and hypothesis were discussed about the results, as well as possible future research directions.1.

Hirsa, A., Osterrieder, J., Hadji-Misheva, B., and Posth, J.-A. (2021). "Deep reinforcement learning on a multi-asset environment for trading." In: *arXiv e-Print*.
Financial trading has been widely analyzed for decades with market participants and academics always looking for advanced methods to improve trading performance. Deep reinforcement learning (DRL), a recently reinvigorated method with significant success in multiple domains, still has to show its benefit in the financial markets. We use a deep Q-network (DQN) to design long-short trading strategies for futures contracts. The state space consists of volatility-normalized daily returns, with buying or selling being the reinforcement learning action and the total reward defined as the cumulative profits from our actions. Our trading strategy is trained and tested both on real and simulated price series and we compare the results with an index benchmark. We analyze how training based on a combination of artificial data and actual price series can be successfully deployed in real markets. The trained reinforcement learning agent is applied to trading the E-mini S&P 500 continuous futures contract. Our results in this study are preliminary and need further improvement.

Hoang, C., Sohn, S., Choi, J., Carvalho, W., and Lee, H. (2021). "Successor Feature Landmarks for Long-Horizon Goal-Conditioned Reinforcement Learning." In: *arXiv e-Print*.
Operating in the real-world often requires agents to learn about a complex environment and apply this understanding to achieve a breadth of goals. This problem, known as goal-conditioned reinforcement learning (GCRL), becomes especially challenging for long-horizon goals. Current methods have tackled this problem by augmenting goal-conditioned policies with graph-based planning algorithms. However, they struggle to scale to large, high-dimensional state spaces and assume access to exploration mechanisms for efficiently collecting training data. In this work, we introduce Successor Feature Landmarks (SFL), a framework for exploring large, high-dimensional environments so as to obtain a policy that is proficient for any goal. SFL leverages the ability of successor features (SF) to capture transition dynamics, using it to drive exploration by estimating state-novelty and to enable high-level planning by abstracting the state-space as a non-parametric landmark-based graph. We further exploit SF to directly compute a goal-conditioned policy for inter-landmark traversal, which we use to execute plans to "frontier" landmarks at the edge of the explored state space. We show in our experiments on MiniGrid and ViZDoom that SFL enables efficient exploration of large, high-dimensional state spaces and outperforms state-of-the-art baselines on long-horizon GCRL tasks.

Hoffman, M., Shahriari, B., Aslanides, J., Barth-Maron, G., Behbahani, F., Norman, T., Abdolmaleki, A., Cassirer, A., Yang, F., Baumli, K., Henderson, S., Novikov, A., Colmenarejo, S. G., Cabi, S., Gulcehre, C., Paine, T. L., Cowie, A., Wang, Z., Piot, B., and Freitas, N. de (2020). "Acme: A Research Framework for Distributed Reinforcement Learning." In: *arXiv e-Print*.
Deep reinforcement learning has led to many recent-and groundbreaking-advancements. However, these advances have often come at the cost of both the scale and complexity of the underlying RL algorithms. Increases in complexity have in turn made it more difficult for researchers to reproduce published RL algorithms or rapidly prototype ideas. To address this, we introduce Acme, a tool to simplify the development of novel RL algorithms that is specifically designed to enable simple agent implementations that can be run at various

scales of execution. Our aim is also to make the results of various RL algorithms developed in academia and industrial labs easier to reproduce and extend. To this end we are releasing baseline implementations of various algorithms, created using our framework. In this work we introduce the major design decisions behind Acme and show how these are used to construct these baselines. We also experiment with these agents at different scales of both complexity and computation-including distributed versions. Ultimately, we show that the design decisions behind Acme lead to agents that can be scaled both up and down and that, for the most part, greater levels of parallelization result in agents with equivalent performance, just faster.

Honchar, A. (2019). *AI for portfolio management: from Markowitz to Reinforcement Learning*. URL: https://medium.com/swlh/ai-for-portfolio-management-from-markowitz-to-reinforcement-learning-cffedcbba566.

The evolution of quantitative asset management techniques with empirical evaluation and Python source code.

Hsu, Y.-C., Lin, H.-W., and Vincent, K. (2017). *Do Cross-Sectional Stock Return Predictors Pass the Test without Data-Snooping Bias?* Tech. rep. Institute of Economics Academia Sinica.

This study examines the possible data-snooping bias as a competing explanation for the anomalies in the cross-section of stock returns. We posit that the exhaustive standalone searches for profitable strategies could lead to recommending spuriously predictive variables. In order to explore the severity of this problem, we use a multiple testing method to evaluate the profitability of portfolios constructed by these predictors. Our empirical analyses suggest that over half of the findings based on individual testing method are no longer statistically significant after we adjust for data-snooping bias. Excluding the micro-cap stocks before portfolios construction and applying the notion of economic significance in this study further weaken the evidence for predictability.

Hsu, P.-H., Han, Q., Wu, W., and Cao, Z. (2018). "Asset allocation strategies, data snooping, and the 1 / N rule." In: *Journal of Banking & Finance* 97, pp. 257–269.

Using a series of advanced tests from White's (2000) Check to correct for data-snooping bias, we assess the out-of-sample performance of various portfolio strategies relative to the naive 1/N rule. When we analyze 16 basic portfolio strategies, 126 learning strategies, and nearly 2,000 extended strategies, we find that some strategies outperform the 1/N rule in conventional tests that do not account for data-snooping bias. However, after we use the new tests that control for such bias, we find that none or very few of these strategies outperform the 1/N rule. Thus, our finding underscores the necessity to control for data-snooping bias when making asset allocation decisions.

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.

Deep reinforcement learning (DRL) is an emerging artificial intelligence (AI) research field which combines deep learning (DL) for policy optimization and reinforcement learning (RL) for goal-oriented self-learning without human intervention. We address major research issues of policy optimization for finance portfolio management. First, we explore one of the deep recurrent neural network (RNN) models, GRUs, to decide the influences of earlier states and actions on policy optimization in non-Markov decision processes. Then, we craft for a viable risk-adjusted reward function to evaluate the expected total rewards for policy. Third, we empower the integration of RL and DL to leverage their respective capabilities to discover an optimal policy. Fourth, we investigate each type of RL approaches for integrating with the DL method while solving the policy optimization problem.

Huang, G., Zhou, X., and Song, Q. (2022). "Deep reinforcement learning for portfolio management." In: *arXiv e-Print*.

In our paper, we apply deep reinforcement learning approach to optimize investment decisions in portfolio management. We make several innovations, such as adding short mechanism and designing an arbitrage mechanism, and applied our model to make decision optimization for several randomly selected portfolios. The experimental results show that our model is able to optimize investment decisions and has the ability to obtain excess return in stock market, and the optimized agent maintains the asset weights at fixed value throughout the trading periods and trades at a very low transaction cost rate. In addition, we redesigned the formula for calculating portfolio asset weights in continuous trading process which can make leverage trading, that fills the theoretical gap in the calculation of portfolio weights when shorting.

Huang, M. and Yu, S. (2020). "A new procedure for resampled portfolio with shrinkaged covariance matrix." In: *Journal of Applied Statistics* 47(44), pp. 642–652.

Dealing with estimation error is an important issue when we implement the mean-variance paradigm for portfolio construction. To tackle the problem, two approaches are proposed in literature, the portfolio resampling technique introduced by Michuad and the well-known shrinkaged covariance matrix method. There are certain evidences on the advantages of shrinkaged covariance over portfolio resampling, however, it is unclear whether a combination of the two approaches could produce a better performance compared with using shrinkaged covariance alone. In this paper, we propose a new algorithm to integrated linear or nonlinear shrinkage estimation with resampled portfolio to achieve a further improvement. Our method are demonstrated via extensive simulation and application in active portfolio management process.

Huang, Z. and Tanaka, F. (2021). "A Modularized and Scalable Multi-Agent Reinforcement Learning-based System for Financial Portfolio Management." In: *arXiv e-Print*.
Financial Portfolio Management is one of the most applicable problems in Reinforcement Learning (RL) by its sequential decision-making nature. Existing RL-based approaches, while inspiring, often lack scalability, reusability, or profundity of intake information to accommodate the ever-changing capital markets. In this paper, we design and develop MSPM, a novel Multi-agent Reinforcement learning-based system with a modularized and scalable architecture for portfolio management. MSPM involves two asynchronously updated units: Evolving Agent Module (EAM) and Strategic Agent Module (SAM). A self-sustained EAM produces signal-comprised information for a specific asset using heterogeneous data inputs, and each EAM possesses its reusability to have connections to multiple SAMs. A SAM is responsible for the assets reallocation of a portfolio using profound information from the EAMs connected. With the elaborate architecture and the multi-step condensation of the volatile market information, MSPM aims to provide a customizable, stable, and dedicated solution to portfolio management that existing approaches do not. We also tackle data-shortage issue of newly-listed stocks by transfer learning, and validate the necessity of EAM. Experiments on 8-year U.S. stock markets data prove the effectiveness of MSPM in profits accumulation by its outperformance over existing benchmarks.

Hulbert, N., Spillers, S., Francis, B., Haines-Temons, J., Romero, K. G., Jager, B. D., Wong, S., Flora, K., Huang, B., and Irissappane, A. A. (2020). "EasyRL: A Simple and Extensible Reinforcement Learning Framework." In: *arXiv e-Print*.
In recent years, Reinforcement Learning (RL), has become a popular field of study as well as a tool for enterprises working on cutting-edge artificial intelligence research. To this end, many researchers have built RL frameworks such as openAI Gym and KerasRL for ease of use. While these works have made great strides towards bringing down the barrier of entry for those new to RL, we propose a much simpler framework called EasyRL, by providing an interactive graphical user interface for users to train and evaluate RL agents. As it is entirely graphical, EasyRL does not require programming knowledge for training and testing simple built-in RL agents. EasyRL also supports custom RL agents and environments, which can be highly beneficial for RL researchers in evaluating and comparing their RL models.

Hwang, I., Xu, S., and In, F. (2018). "Naive versus optimal diversification: Tail risk and performance." In: *European Journal of Operational Research* 265(1), pp. 372–388.
It is well documented in portfolio optimization that naive diversification outperforms optimal mean-variance diversification because the latter is subject to severe estimation error. Our study provides an alternative explanation for the outperformance of naive diversification by examining the tail risk of naive diversification relative to optimal mean-variance diversification. We utilize a rolling-sample approach and compare the out-of-sample performance and tail risk of various optimal strategies to that of the naive diversification strategy. Using portfolios consisting of individual stocks, we show that for portfolios containing relatively small number of stocks, naive diversification outperforms optimal mean-variance diversification and is less exposed to tail risk. However, for relatively large number of stocks in the portfolio, naive diversification maintains its superior performance but increases tail risk and results in more concave portfolio returns. These results imply that the outperformance of naive diversification acts as compensation for the increase in tail risk and concavity.

Ielpo, F., Merhy, C., and Simon, G. (2017). *Engineering Investment Process: Making Value Creation Repeatable*. Elsevier. 430 pp.
The book explores the quantitative steps of a financial investment process. The authors study how these steps are articulated in order to make any value creation, whatever the asset class, consistent and robust. The discussion includes factors, portfolio allocation, statistical and economic backtesting, but also the influence of negative rates, dynamical trading, state-space models, stylized facts, liquidity issues, or data biases. Besides

the quantitative concepts detailed here, the reader will find useful references to other works to develop an in-depth understanding of an investment process.

Irlam, G. (2020a). "Machine learning for retirement planning." In: *The Journal of Retirement* 8(1), pp. 32–29.
Machine learning provides a new approach to solving problems in many fields. This article explores the use of machine learning to solve the retirement portfolio problem: deciding how much wealth to consume and how to allocate the remainder. After first reviewing existing approaches to the portfolio problem, this article looks in detail at the use of reinforcement learning. For simple financial scenarios where the optimal solution is known, reinforcement learning is found to deliver to within a few percent of the optimal solution. For more complicated financial scenarios, with no known optimal solution, reinforcement learning outperforms other common approaches. Reinforcement learning proves capable of optimizing highly complex financial models, including the effects of income taxes, mean-reverting asset classes, and time-varying bond yield curves, all of which other approaches cannot handle. Reinforcement learning appears to be the first fundamentally new approach to the portfolio problem in over 50 years.

Irlam, G. (2020b). "Multi Scenario Financial Planning via Deep Reinforcement Learning AI." In: *SSRN e-Print*.
Financial planning via deep reinforcement learning holds much promise. One implementation, AIPlanner, delivered near optimal financial results, but had a major shortcoming. It required training a separate neural network model for each financial scenario. This paper describes extending AIPlanner so that a small family of trained neural network models are capable of rapidly producing financial plans for a wide variety of financial scenarios. Additionally AIPlanner is extended to produce results over the lifecycle, both pre and post retirement, and for couples, as well as individuals. A reasonably realistic income tax model is incorporated. And finally, a more realistic stock model is used. Over the lifecycle, compared to the best discovered alternative strategy, reinforcement learning was found to effectively deliver 14% more retirement consumption.

Ivanov, S. and D'yakonov, A. (2019). "Modern Deep Reinforcement Learning Algorithms." In: *arXiv e-Print*.
Recent advances in Reinforcement Learning, grounded on combining classical theoretical results with Deep Learning paradigm, led to breakthroughs in many artificial intelligence tasks and gave birth to Deep Reinforcement Learning (DRL) as a field of research. In this work latest DRL algorithms are reviewed with a focus on their theoretical justification, practical limitations and observed empirical properties.

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*.
In this paper, we construct a pipeline to investigate heuristic diversification strategies in asset allocation. We use machine learning concepts ("explainable AI") to compare the robustness of different strategies and back out implicit rules for decision making.In a first step, we augment the asset universe (the empirical dataset) with a range of scenarios generated with a block bootstrap from the empirical dataset.Second, we backtest the candidate strategies over a long period of time, checking their performance variability. Third, we use XGBoost as a regression model to connect the difference between the measured performances between two strategies to a pool of statistical features of the portfolio universe tailored to the investigated strategy. Finally, we employ the concept of Shapley values to extract the relationships that the model could identify between the portfolio characteristics and the statistical properties of the asset universe. We test this pipeline for studying risk-parity strategies with a volatility target, and in particular, comparing the machine learning-driven Hierarchical Risk Parity (HRP) to the classical Equal Risk Contribution (ERC) strategy.In the augmented dataset built from a multi-asset investment universe of commodities, equities and fixed income futures, we find that HRP better matches the volatility target, and shows better risk-adjusted performances. Finally, we train XGBoost to learn the difference between the realized Calmar ratios of HRP and ERC and extract explanations.The explanations provide fruitful ex-post indications of the connection between the statistical properties of the universe and the strategy performance in the training set. For example, the model confirms that features addressing the hierarchical properties of the universe are connected to the relative performance of HRP respect to ERC.

Jaimungal, S. (2022). "Reinforcement learning and stochastic optimisation." In: *Finance and Stochastics* 26(1), pp. 103–129.
At the heart of financial mathematics lie stochastic optimisation problems. Traditional approaches to solving such problems, while applicable to broad classes of models, require specifying a model to complete the analysis and obtain implementable results. Even then, the curse of dimensionality challenges the viability of conventional methods to settings of practical relevance. In contrast, machine learning, and reinforcement learning (RL) particularly, promises to learn from data and overcome the curse of dimensionality simultane-

ously. This article touches on several approaches in the extant literature that are well positioned to merge our traditional techniques with RL.

Jaimungal, S., Pesenti, S. M., Wang, Y. S., and Tatsat, H. (2021). "Robust Risk-Aware Reinforcement Learning." In: *SSRN e-Print*.

We present a reinforcement learning (RL) approach for robust optimisation of risk-aware performance criteria. To allow agents to express a wide variety of risk-reward profiles, we assess the value of a policy using rank dependent expected utility (RDEU). RDEU allows the agent to seek gains, while simultaneously protecting themselves against downside events. To robustify optimal policies against model uncertainty, we assess a policy not by its distribution, but rather, by the worst possible distribution that lies within a Wasserstein ball around it. Thus, our problem formulation may be viewed as an actor choosing a policy (the outer problem), and the adversary then acting to worsen the performance of that strategy (the inner problem). We develop explicit policy gradient formulae for the inner and outer problems, and show its efficacy on three prototypical financial problems: robust portfolio allocation, optimising a benchmark, and statistical arbitrage. Code at `https://github.com/sebjai/robust-risk-aware-rl`.

Jaisson, T. (2022). "Deep differentiable reinforcement learning and optimal trading." In: *arXiv e-Print*.

In many reinforcement learning applications, the underlying environment reward and transition functions are explicitly known differentiable functions. This enables us to use recent research which applies machine learning tools to stochastic control to find optimal action functions. In this paper, we define differentiable reinforcement learning as a particular case of this research. We find that incorporating deep learning in this framework leads to more accurate and stable solutions than those obtained from more generic actor critic algorithms. We apply this deep differentiable reinforcement learning (DDRL) algorithm to the problem of one asset optimal trading strategies in various environments where the market dynamics are known. Thanks to the stability of this method, we are able to efficiently find optimal strategies for complex multi-scale market models. We also extend these methods to simultaneously find optimal action functions for a wide range of environment parameters. This makes it applicable to real life financial signals and portfolio optimization where the expected return has multiple time scales. In the case of a slow and a fast alpha signal, we find that the optimal trading strategy consists in using the fast signal to time the trades associated to the slow signal.

Janner, M., Li, Q., and Levine, S. (2021). "Reinforcement Learning as One Big Sequence Modeling Problem." In: *arXiv e-Print*.

Reinforcement learning (RL) is typically concerned with estimating single-step policies or single-step models, leveraging the Markov property to factorize the problem in time. However, we can also view RL as a sequence modeling problem, with the goal being to predict a sequence of actions that leads to a sequence of high rewards. Viewed in this way, it is tempting to consider whether powerful, high-capacity sequence prediction models that work well in other domains, such as natural-language processing, can also provide simple and effective solutions to the RL problem. To this end, we explore how RL can be reframed as "one big sequence modeling" problem, using state-of-the-art Transformer architectures to model distributions over sequences of states, actions, and rewards. Addressing RL as a sequence modeling problem significantly simplifies a range of design decisions: we no longer require separate behavior policy constraints, as is common in prior work on offline model-free RL, and we no longer require ensembles or other epistemic uncertainty estimators, as is common in prior work on model-based RL. All of these roles are filled by the same Transformer sequence model. In our experiments, we demonstrate the flexibility of this approach across long-horizon dynamics prediction, imitation learning, goal-conditioned RL, and offline RL.

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.

The authors propose the use of goal based investing- or private ALM, as they prefer to call it- to tailor a dynamic investment strategy to the needs of individual clients. They argue that this approach is superior to the - one-size-fits-all,- target-date-oriented static allocation path used in most current life cycle funds. They present the two pillars of their approach: the methodology for obtaining financial and economic scenarios, and the methodology of the goal-oriented dynamic allocation strategy. This approach reduces the risks and improves the feasibility of meeting the clients' goals.

Jordan, S. M., Chandak, Y., Cohen, D., Zhang, M., and Thomas, P. (2020). "Evaluating the Performance of Reinforcement Learning Algorithms." In: *Thirty-seventh International Conference on Machine Learning ICML*.

Performance evaluations are critical for quantifying algorithmic advances in reinforcement learning. Recent reproducibility analyses have shown that reported performance results are often inconsistent and difficult to replicate. In this work, we argue that the inconsistency of performance stems from the use of flawed evaluation metrics. Taking a step towards ensuring that reported results are consistent, we propose a new comprehensive evaluation methodology for reinforcement learning algorithms that produces reliable measurements of performance both on a single environment and when aggregated across environments. We demonstrate this method by evaluating a broad class of reinforcement learning algorithms on standard benchmark tasks.

Katongo, M. and Bhattacharyya, R. (2021). "The Use of Deep Reinforcement Learning in Tactical Asset Allocation." In: *SSRN e-Print*.
The Tactical Asset Allocation (TAA) problem is a problem to accurately capture short to medium term market trends and anomalies in order to allocate the assets in a portfolio so as to optimize its performance by increasing the risk adjusted returns. This project seeks to address the Tactical Asset Allocation problem by employing Deep Reinforcement Learning (DRL) Algorithms in a Machine Learning Environment as well as employing Neural Network Autoencoders for selection of portfolio assets. This paper presents the implementation of this proposed methodology applied to 30 stocks of the Dow Jones Industrial Average (DJIA). In (1), the Introduction to the project objectives is done with the Problem Description presented in (2). Part (3) presents the literature review of similar studies in the subject area. The methodology used for our implementation is presented in (4) whilst (5) and (6) presents the benchmark portfolios and the DRL portfolios development respectively. The evaluation of the performance of the models is presented in (7) and we present our conclusions and the future works in (8).

Kazak, E. and Pohlmeier, W. (2019). "Testing out-of-sample portfolio performance." In: *International Journal of Forecasting* 35(2), pp. 540–554.
This paper studies the quality of portfolio performance tests based on out-of-sample returns. By disentangling the components of the out-of-sample performance, we show that the observed differences are driven largely by the differences in estimation risk. Our Monte Carlo study reveals that the puzzling empirical findings of inferior performances of theoretically superior strategies result mainly from the low power of these tests. Thus, our results provide an explanation as to why the null hypothesis of equal performance of the simple equally-weighted portfolio compared to many theoretically-superior alternative strategies cannot be rejected in many out-of-sample horse races. Our findings turn out to be robust with respect to different designs and the implementation strategies of the tests. For the applied researcher, we provide some guidance as to how to cope with the problem of low power. In particular, we make use of a novel pretest-based portfolio strategy to show how the information regarding performance tests can be used optimally.

Kazak, E. and Pohlmeier, W. (2020). *Portfolio Pretesting with Machine Learning*. Tech. rep. University of Lancaster.
This paper exploits the idea of pretesting to choose between competing portfolio strategies. We propose an estimator for a portfolio weight vector, which optimally trades off between Type I and Type II errors when choosing the best investment strategy. Furthermore we accommodate the idea of bagging in the portfolio testing problems, which helps to avoid sharp thresholding and reduces the amount of portfolio turnover. Our approach borrows from both shrinkage and forecast combination literature. The portfolio weights of our strategy are weighted averages of the portfolio weights from a set of stand-alone strategies. More specifically, the weights are generated from a pseudo out-of-sample portfolio pretesting, such that they reflect the probability that a given strategy will be overall best performing. Contrary to previous approaches, the shrinkage intensity is continuously updated to incorporate the most recent information in the rolling window forecasting set-up. We show that the bagged pretest estimator performs exceptionally well, especially when combined with adaptive smoothing. The resulting strategy allows for a flexible and smooth switch between the underlying strategies and is shown to outperform the corresponding stand-alone strategies.

Khetarpal, K., Riemer, M., Rish, I., and Precup, D. (2021). "Towards Continual Reinforcement Learning: A Review and Perspectives." In: *arXiv e-Print*.
In this article, we aim to provide a literature review of different formulations and approaches to continual reinforcement learning (RL), also known as lifelong or non-stationary RL. We begin by discussing our perspective on why RL is a natural fit for studying continual learning. We then provide a taxonomy of different continual RL formulations and mathematically characterize the non-stationary dynamics of each setting. We go on to discuss evaluation of continual RL agents, providing an overview of benchmarks used in the literature and important metrics for understanding agent performance. Finally, we highlight open problems

and challenges in bridging the gap between the current state of continual RL and findings in neuroscience. While still in its early days, the study of continual RL has the promise to develop better incremental reinforcement learners that can function in increasingly realistic applications where non-stationarity plays a vital role. These include applications such as those in the fields of healthcare, education, logistics, and robotics.

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.

In this paper, we propose a goal-based investment model that is suitable for personalized wealth management. The model only requires a few intuitive inputs such as size of wealth, investment amount, and consumption goals from individual investors. In particular, a priority level can be assigned to each consumption goal and the model provides a holistic solution based on a sequential approach starting with the highest priority. This allows strict prioritization by maximizing the probability of achieving higher priority goals that are not affected by goals with lower priorities. Furthermore, the proposed model is formulated as a linear program that efficiently finds the optimal financial plan. With its simplicity, flexibility, and computational efficiency, the proposed goal-based investment model provides a new framework for automated investment management services.

Kirk, R., Zhang, A., Grefenstette, E., and Rocktaschel, T. (2022). "A Survey of Generalisation in Deep Reinforcement Learning." In: *arXiv e-Print*.

The study of generalisation in deep Reinforcement Learning (RL) aims to produce RL algorithms whose policies generalise well to novel unseen situations at deployment time, avoiding overfitting to their training environments. Tackling this is vital if we are to deploy reinforcement learning algorithms in real world scenarios, where the environment will be diverse, dynamic and unpredictable. This survey is an overview of this nascent field. We provide a unifying formalism and terminology for discussing different generalisation problems, building upon previous works. We go on to categorise existing benchmarks for generalisation, as well as current methods for tackling the generalisation problem. Finally, we provide a critical discussion of the current state of the field, including recommendations for future work. Among other conclusions, we argue that taking a purely procedural content generation approach to benchmark design is not conducive to progress in generalisation, we suggest fast online adaptation and tackling RL-specific problems as some areas for future work on methods for generalisation, and we recommend building benchmarks in underexplored problem settings such as offline RL generalisation and reward-function variation.

Kolesnikov, S. and Hrinchuk, O. (2019). "Catalyst.RL: A Distributed Framework for Reproducible RL Research." In: *arXiv e-Print*.

Despite the recent progress in deep reinforcement learning field (RL), and, arguably because of it, a large body of work remains to be done in reproducing and carefully comparing different RL algorithms. We present catalyst.RL, an open source framework for RL research with a focus on reproducibility and flexibility. Main features of our library include large-scale asynchronous distributed training, easy-to-use configuration files with the complete list of hyperparameters for the particular experiments, efficient implementations of various RL algorithms and auxiliary tricks, such as frame stacking, n-step returns, value distributions, etc. To vindicate the usefulness of our framework, we evaluate it on a range of benchmarks in a continuous control, as well as on the task of developing a controller to enable a physiologically-based human model with a prosthetic leg to walk and run. The latter task was introduced at NeurIPS 2018 AI for Prosthetics Challenge, where our team took the 3rd place, capitalizing on the ability of catalyst.RL to train high-quality and sample-efficient RL agents.

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

We give an overview and outlook of the field of reinforcement learning as it applies to solving financial applications of intertemporal choice. In finance, common problems of this kind include pricing and hedging of contingent claims, investment and portfolio allocation, buying and selling a portfolio of securities subject to transaction costs, market making, asset liability management and optimization of tax consequences, to name a few. Reinforcement learning allows us to solve these dynamic optimization problems in an almost model-free way, relaxing the assumptions often needed for classical approaches.A main contribution of this article is the elucidation of the link between these dynamic optimization problem and reinforcement learning, concretely addressing how to formulate expected intertemporal utility maximization problems using modern machine learning techniques.

Krishnan, S., Garg, A., Liaw, R., Miller, L., Pokorny, F. T., and Goldberg, K. (2016). "HIRL: Hierarchical Inverse Reinforcement Learning for Long-Horizon Tasks with Delayed Rewards." In: *arXiv e-Print*.
Reinforcement Learning (RL) struggles in problems with delayed rewards, and one approach is to segment the task into sub-tasks with incremental rewards. We propose a framework called Hierarchical Inverse Reinforcement Learning (HIRL), which is a model for learning sub-task structure from demonstrations. HIRL decomposes the task into sub-tasks based on transitions that are consistent across demonstrations. These transitions are defined as changes in local linearity w.r.t to a kernel function. Then, HIRL uses the inferred structure to learn reward functions local to the sub-tasks but also handle any global dependencies such as sequentiality. We have evaluated HIRL on several standard RL benchmarks: Parallel Parking with noisy dynamics, Two-Link Pendulum, 2D Noisy Motion Planning, and a Pinball environment. In the parallel parking task, we find that rewards constructed with HIRL converge to a policy with an 80% success rate in 32% fewer time-steps than those constructed with Maximum Entropy Inverse RL (MaxEnt IRL), and with partial state observation, the policies learned with IRL fail to achieve this accuracy while HIRL still converges. We further find that that the rewards learned with HIRL are robust to environment noise where they can tolerate 1 stdev. of random perturbation in the poses in the environment obstacles while maintaining roughly the same convergence rate. We find that HIRL rewards can converge up-to 6x faster than rewards constructed with IRL.

Kuntz, L.-C. (2018). "Portfolio Strategies with Classical and Alternative Benchmarks." PhD thesis. Georg August University of Gottingen.
This dissertation addresses different key elements in portfolio management. It intends to improve and analyze influences on portfolio strategies and their performance. Likewise, it aims at the systematization and extension of benchmark specifications as well as their effect on portfolio strategies. Each chapter focuses on a different aspect of developing and implementing portfolio strategies. The dissertation seeks to contribute to the advancement of portfolio strategies by making the performance generating process and influences on it more comprehensible and transparent. In doing so, it attempts to strengthen the awareness of the impact of the exact design of portfolio strategies and benchmarks on the resulting portfolio and its performance. The key findings of this dissertation can be summarized as follows: The benchmark specification, especially in terms of the investible universe and the inherent risk conception, has substantial influence on the explicit design and performance of portfolio strategies. In general, the specification of the benchmark and design of portfolio strategies should be carefully considered and the implementation should be well thought out. Alternative risk conceptions, such as regret risk, can be applied to portfolio selection and lead to clearly different portfolio compositions. Moreover, timing strategies can be improved by choosing a careful investment approach on the basis of distributional regressions. All empirical work 3 of this thesis has in common that it pursues different ideas to set up portfolio strategies while explicitly addressing the benchmark specification used for the implementation and evaluation of said strategies.

Kuttler, H., Nardelli, N., Lavril, T., Selvatici, M., Sivakumar, V., Rocktaschel, T., and Grefenstette, E. (2019). "TorchBeast: A PyTorch Platform for Distributed RL." In: *arXiv e-Print*.
TorchBeast is a platform for reinforcement learning (RL) research in PyTorch. It implements a version of the popular IMPALA algorithm for fast, asynchronous, parallel training of RL agents. Additionally, TorchBeast has simplicity as an explicit design goal: We provide both a pure-Python implementation ("MonoBeast") as well as a multi-machine high-performance version ("PolyBeast"). In the latter, parts of the implementation are written in C++, but all parts pertaining to machine learning are kept in simple Python using PyTorch, with the environments provided using the OpenAI Gym interface. This enables researchers to conduct scalable RL research using TorchBeast without any programming knowledge beyond Python and PyTorch. In this paper, we describe the TorchBeast design principles and implementation and demonstrate that it performs on-par with IMPALA on Atari. TorchBeast is released as an open-source package under the Apache 2.0 license and is available at this https URL.

Lambert, N. O., Wilcox, A., Zhang, H., Pister, K. S. J., and Calandra, R. (2021). "Learning Accurate Long-term Dynamics for Model-based Reinforcement Learning." In: *arXiv e-Print*.
Accurately predicting the dynamics of robotic systems is crucial for model-based control and reinforcement learning. The most common way to estimate dynamics is by fitting a one-step ahead prediction model and using it to recursively propagate the predicted state distribution over long horizons. Unfortunately, this approach is known to compound even small prediction errors, making long-term predictions inaccurate. In this paper, we propose a new parametrization to supervised learning on state-action data to stably predict

at longer horizons – that we call a trajectory-based model. This trajectory-based model takes an initial state, a future time index, and control parameters as inputs, and directly predicts the state at the future time index. Experimental results in simulated and real-world robotic tasks show that trajectory-based models yield significantly more accurate long term predictions, improved sample efficiency, and the ability to predict task reward. With these improved prediction properties, we conclude with a demonstration of methods for using the trajectory-based model for control.

Lapan, M. (2020). *Deep Reinforcement Learning Hands-On*. 2nd ed. Packt. 826 pp.
New edition of the bestselling guide to deep reinforcement learning and how it's used to solve complex real-world problems. Revised and expanded to include multi-agent methods, discrete optimization, RL in robotics, advanced exploration techniques, and more.

Laskin, M., Lee, K., Stooke, A., Pinto, L., Abbeel, P., and Srinivas, A. (2020). "Reinforcement Learning with Augmented Data." In: *arXiv e-Print*.
Learning from visual observations is a fundamental yet challenging problem in reinforcement learning (RL). Although algorithmic advancements combined with convolutional neural networks have proved to be a recipe for success, current methods are still lacking on two fronts: (a) sample efficiency of learning and (b) generalization to new environments. To this end, we present RAD: Reinforcement Learning with Augmented Data, a simple plug-and-play module that can enhance any RL algorithm. We show that data augmentations such as random crop, color jitter, patch cutout, and random convolutions can enable simple RL algorithms to match and even outperform complex state-of-the-art methods across common benchmarks in terms of data-efficiency, generalization, and wall-clock speed. We find that data diversity alone can make agents focus on meaningful information from high-dimensional observations without any changes to the reinforcement learning method. On the DeepMind Control Suite, we show that RAD is state-of-the-art in terms of data-efficiency and performance across 15 environments. We further demonstrate that RAD can significantly improve the test-time generalization on several OpenAI ProcGen benchmarks. Finally, our customized data augmentation modules enable faster wall-clock speed compared to competing RL techniques. Our RAD module and training code are available at `https://github.com/MishaLaskin/rad`.

Laskin, M., Yarats, D., Liu, H., Lee, K., Zhan, A., Lu, K., Cang, C., Pinto, L., and Abbeel, P. (2021). "URLB: Unsupervised Reinforcement Learning Benchmark." In: *arXiv e-Print*.
Deep Reinforcement Learning (RL) has emerged as a powerful paradigm to solve a range of complex yet specific control tasks. Yet training generalist agents that can quickly adapt to new tasks remains an outstanding challenge. Recent advances in unsupervised RL have shown that pre-training RL agents with self-supervised intrinsic rewards can result in efficient adaptation. However, these algorithms have been hard to compare and develop due to the lack of a unified benchmark. To this end, we introduce the Unsupervised Reinforcement Learning Benchmark (URLB). URLB consists of two phases: reward-free pre-training and downstream task adaptation with extrinsic rewards. Building on the DeepMind Control Suite, we provide twelve continuous control tasks from three domains for evaluation and open-source code for eight leading unsupervised RL methods. We find that the implemented baselines make progress but are not able to solve URLB and propose directions for future research.

Lazaridis, A., Fachantidis, A., and Vlahavas, I. (2020). "Deep Reinforcement Learning: A State-of-the-Art Walkthrough." In: *Journal of Artificial Intelligence Research* 69, pp. 1421–1471.
Deep Reinforcement Learning is a topic that has gained a lot of attention recently, due to the unprecedented achievements and remarkable performance of such algorithms in various benchmark tests and environmental setups. The power of such methods comes from the combination of an already established and strong field of Deep Learning, with the unique nature of Reinforcement Learning methods. It is, however, deemed necessary to provide a compact, accurate and comparable view of these methods and their results for the means of gaining valuable technical and practical insights. In this work we gather the essential methods related to Deep Reinforcement Learning, extracting common property structures for three complementary core categories: a) Model-Free, b) Model-Based and c) Modular algorithms. For each category, we present, analyze and compare state-of-the-art Deep Reinforcement Learning algorithms that achieve high performance in various environments and tackle challenging problems in complex and demanding tasks. In order to give a compact and practical overview of their differences, we present comprehensive comparison figures and tables, produced by reported performances of the algorithms under two popular simulation platforms: the Atari Learning Environment and the MuJoCo physics simulation platform. We discuss the key differences of the various kinds

of algorithms, indicate their potential and limitations, as well as provide insights to researchers regarding future directions of the field.

Lehnert, L. and Littman, M. L. (2020). "Successor Features Combine Elements of Model-Free and Model-based Reinforcement Learning." In: *Journal of Machine Learning Research* 21(196138), pp. 1–53.
A key question in reinforcement learning is how an intelligent agent can generalize knowledge across different inputs. By generalizing across different inputs, information learned for one input can be immediately reused for improving predictions for another input. Reusing information allows an agent to compute an optimal decision-making strategy using less data. State representation is a key element of the generalization process, compressing a high-dimensional input space into a low-dimensional latent state space. This article analyzes properties of different latent state spaces, leading to new connections between model-based and model-free reinforcement learning. Successor features, which predict frequencies of future observations, form a link between model-based and model-free learning: Learning to predict future expected reward outcomes, a key characteristic of model-based agents, is equivalent to learning successor features. Learning successor features is a form of temporal difference learning and is equivalent to learning to predict a single policy's utility, which is a characteristic of model-free agents. Drawing on the connection between model-based reinforcement learning and successor features, we demonstrate that representations that are predictive of future reward outcomes generalize across variations in both transitions and rewards. This result extends previous work on successor features, which is constrained to fixed transitions and assumes re-learning of the transferred state representation.

Li, B., Cowen-Rivers, A., Kozakowski, P., Tao, D., Kamalakara, S., Rajkumar, N., Sezhiyan, H., Huang, S., and Gomez, A. (2019). "RL: Generic reinforcement learning codebase in TensorFlow." In: *The Journal of Open Source Software* 4(42), p. 1524.
Vast reinforcement learning (RL) research groups, such as DeepMind and OpenAI, have theirinternal (private) reinforcement learning codebases, which enable quick prototyping and com-paring of ideas to many state-of-the-art (SOTA) methods. We argue the five fundamentalproperties of a sophisticated research codebase are: modularity, reproducibility, many RL al-gorithms pre-implemented, speed and ease of running on different hardware/ integration withvisualization packages. Currently, there does not exist any RL codebase, to the author'sknowledge, which contains all the five properties, particularly with TensorBoard logging andabstracting away cloud hardware such as TPU's from the user. The codebase aims to helpdistil the best research practices into the community as well as ease the entry access andaccelerate the pace of the field. More detailed documentation can be found here.

Li, L. (2021a). "An Automated Portfolio Trading System with Feature Preprocessing and Recurrent Reinforcement Learning." In: *arXiv e-Print*.
We propose a novel portfolio trading system, which contains a feature preprocessing module and a trading module. The feature preprocessing module consists of various data processing operations, while in the trading part, we integrate the portfolio weight rebalance function with the trading algorithm and make the trading system fully automated and suitable for individual investors, holding a handful of stocks. The data preprocessing procedures are applied to remove the white noise in the raw data set and uncover the general pattern underlying the data set before the processed feature set is inputted into the trading algorithm. Our empirical results reveal that the proposed portfolio trading system can efficiently earn high profit and maintain a relatively low drawdown, which clearly outperforms other portfolio trading strategies.

Li, L. (2021b). "Financial Trading with Feature Preprocessing and Recurrent Reinforcement Learning." In: *arXiv e-Print*.
Financial trading aims to build profitable strategies to make wise investment decisions in the financial market. It has attracted interests in the machine learning community for a long time. This paper proposes to trade financial assets automatically using feature preprocessing skills and Recurrent Reinforcement Learning (RRL) algorithm. The strategy starts from technical indicators extracted from assets' market information. Then these technical indicators are preprocessed by Principal Component Analysis (PCA) and Discrete Wavelet Transform (DWT) and eventually inputted to the RRL algorithm to do the trading. The extensive empirical evidence shows that the proposed strategy is not only effective and robust in its performance, but also can mitigate the drawbacks underlying the initial trading using RRL.

Li, R., Cai, Z., Huang, T., and Zhu, W. (2021a). "Anchor: The achieved goal to replace the subgoal for hierarchical reinforcement learning." In: *Knowledge-Based Systems* 225, p. 107128.

Hierarchical reinforcement learning (HRL) extends traditional reinforcement learning methods to complex tasks, such as the continuous control task with long horizon. As an effective paradigm for HRL, the subgoal-based HRL method uses subgoals to provide intrinsic motivation which helps the agent to reach the desired goal. However, it is tough to determine the subgoal. In this paper, we present a new concept called anchor to replace the subgoal. Our anchor is selected from the achieved goals of the agent. By the anchor, we propose a new HRL method which encourages the agent to move fast away from the corresponding anchor in the right direction of reaching the desired goal. Specifically, for moving fast, our new method uses an intrinsic reward computed by the distance between the current achieved goal and the corresponding anchor. Meanwhile, for moving in the right direction, it weights the intrinsic reward by the extrinsic rewards collected in the process of moving away from the corresponding anchor. The experiments demonstrate the effectiveness of the proposed method on the continuous control task with long horizon.

Li, S., Zang, Z., Wu, D., Chen, Z., and Li, S. Z. (2021b). "GenURL: A General Framework for Unsupervised Representation Learning." In: *arXiv e-Print*.

Recently unsupervised representation learning (URL) has achieved remarkable progress in various scenarios. However, most methods are specifically designed based on specific data characters or task assumptions. Based on the manifold assumption, we regard most URL problems as an embedding problem that seeks an optimal low-dimensional representation of the given high-dimensional data. We split the embedding process into two steps, data structural modeling and low-dimensional embedding, and propose a general similarity-based framework called GenURL. Specifically, we provide a general method to model data structures by adaptively combining graph distances on the feature space and predefined graphs, then propose robust loss functions to learn the low-dimensional embedding. Combining with a specific pretext task, we can adapt GenURL to various URL tasks in a unified manner and achieve state-of-the-art performance, including self-supervised visual representation learning, unsupervised knowledge distillation, graph embeddings, and dimension reduction. Moreover, ablation studies of loss functions and basic hyper-parameter settings in GenURL illustrate the data characters of various tasks.

Li, Z., Liu, X.-Y., Zheng, J., Wang, Z., Walid, A., and Guo, J. (2021c). "FinRL-Podracer: High Performance and Scalable Deep Reinforcement Learning for Quantitative Finance." In: *ACM International Conference on AI in Finance*.

Machine learning techniques are playing more and more important roles in finance market investment. However, finance quantitative modeling with conventional supervised learning approaches has a number of limitations. The development of deep reinforcement learning techniques is partially addressing these issues. Unfortunately, the steep learning curve and the difficulty in quick modeling and agile development are impeding finance researchers from using deep reinforcement learning in quantitative trading. In this paper, we propose an RLOps in finance paradigm and present a FinRL-Podracer framework to accelerate the development pipeline of deep reinforcement learning (DRL)-driven trading strategy and to improve both trading performance and training efficiency. FinRL-Podracer is a cloud solution that features high performance and high scalability and promises continuous training, continuous integration, and continuous delivery of DRL-driven trading strategies, facilitating a rapid transformation from algorithmic innovations into a profitable trading strategy. First, we propose a generational evolution mechanism with an ensemble strategy to improve the trading performance of a DRL agent, and schedule the training of a DRL algorithm onto a GPU cloud via multi-level mapping. Then, we carry out the training of DRL components with high-performance optimizations on GPUs. Finally, we evaluate the FinRL-Podracer framework for a stock trend prediction task on an NVIDIA DGX SuperPOD cloud. FinRL-Podracer outperforms three popular DRL libraries Ray RLlib, Stable Baseline 3 and FinRL, i.e., 12% ∼35% improvements in annual return, 0.1 ∼0.6 improvements in Sharpe ratio and 3 times ∼7 times speed-up in training time. We show the high scalability by training a trading agent in 10 minutes with 80 A100 GPUs, on NASDAQ-100 constituent stocks with minute-level data over 10 years.

Liao, S.-L., Lin, S.-K., Kuang, X.-J., and Chen, T. (2022). "Portfolio Allocation with Dynamic Risk Preference Via Reinforcement Learning: Evidence from the Taiwan 50 Index." In: *SSRN e-Print*.

In reinforcement learning, an agent interacts with its environment to act according to a policy in accordance with the observation state; these are executed based on a class of neural networks. In the interaction between states and actions, the series of decision sequences are often described using Markovian decision processes, and the corresponding reward is calculated before the appropriate parameters of the strategy are identified by maximizing the advantage function. In this study, we construct the characteristics of an investment

environment based on market data in an enhanced learning framework and identified portfolio allocation conditions corresponding to different risk preferences by using proximal policy optimization with the long short-term memory framework. For a given risk preference, a portfolio weighting that maximizes expected returns or minimizes risk can be identified on the efficiency frontier in conjunction with Markowitz portfolio theory. In our empirical study, we construct portfolios from specific stocks in the Taiwan stock market and then train and test models with a moving window to compare the performance and investment risk of the reinforcement learning dynamic portfolio allocation model against other models under various market conditions.

Liu, D., Alnegheimish, S., Zytek, A., and Veeramachaneni, K. (2021a). "MTV: Visual Analytics for Detecting, Investigating, and Annotating Anomalies in Multivariate Time Series." In: *arXiv e-Print*.
Detecting anomalies in time-varying multivariate data is crucial in various industries for the predictive maintenance of equipment. Numerous machine learning (ML) algorithms have been proposed to support automated anomaly identification. However, a significant amount of human knowledge is still required to interpret, analyze, and calibrate the results of automated analysis. This paper investigates current practices used to detect and investigate anomalies in time series data in industrial contexts and identifies corresponding needs. Through iterative design and working with nine experts from two industry domains (aerospace and energy), we characterize six design elements required for a successful visualization system that supports effective detection, investigation, and annotation of time series anomalies. We summarize an ideal human-AI collaboration workflow that streamlines the process and supports efficient and collaborative analysis. We introduce MTV (Multivariate Time Series Visualization), a visual analytics system to support such workflow. The system incorporates a set of novel visualization and interaction designs to support multi-faceted time series exploration, efficient in-situ anomaly annotation, and insight communication. Two user studies, one with 6 spacecraft experts (with routine anomaly analysis tasks) and one with 25 general end-users (without such tasks), are conducted to demonstrate the effectiveness and usefulness of MTV.

Liu, M., Zhu, M., and Zhang, W. (2022). "Goal-Conditioned Reinforcement Learning: Problems and Solutions." In: *arXiv e-Print*.
Goal-conditioned reinforcement learning (GCRL), related to a set of complex RL problems, trains an agent to achieve different goals under particular scenarios. Compared to the standard RL solutions that learn a policy solely depending on the states or observations, GCRL additionally requires the agent to make decisions according to different goals. In this survey, we provide a comprehensive overview of the challenges and algorithms for GCRL. Firstly, we answer what the basic problems are studied in this field. Then, we explain how goals are represented and present how existing solutions are designed from different points of view. Finally, we make the conclusion and discuss potential future prospects that recent researches focus on.

Liu, X.-Y., Yang, H., Gao, J., and Wang, C. (2021b). "FinRL: Deep Reinforcement Learning Framework to Automate Trading in Quantitative Finance." In: *SSRN e-Print*.
Deep reinforcement learning (DRL) has been envisioned to have a competitive edge in quantitative finance. However, there is a steep development curve for quantitative traders to obtain an agent that automatically positions to win in the market, namely to decide where to trade, at what price and what quantity, due to the error-prone programming and arduous debugging. In this paper, we present the first open-source framework FinRL as a full pipeline to help quantitative traders overcome the steep learning curve. FinRL is featured with simplicity, applicability and extensibility under the key principles, full-stack framework, customization, reproducibility and hands-on tutoring. Embodied as a three-layer architecture with modular structures, FinRL implements fine-tuned state-of-the-art DRL algorithms and common reward functions, while alleviating the debugging work- loads. Thus, we help users pipeline the strategy design at a high turnover rate. At multiple levels of time granularity, FinRL simu- lates various markets as training environments using historical data and live trading APIs. Being highly extensible, FinRL reserves a set of user-import interfaces and incorporates trading constraints such as market friction, market liquidity and investor's risk-aversion. Moreover, serving as practitioners' stepping stones, typical trad- ing tasks are provided as step-by-step tutorials, e.g., stock trading, portfolio allocation, cryptocurrency trading, etc.

Llorente, F., Martino, L., Read, J., and Delgado, D. (2021). "A survey of Monte Carlo methods for noisy and costly densities with application to reinforcement learning." In: *arXiv e-Print*.
This survey gives an overview of Monte Carlo methodologies using surrogate models, for dealing with densities which are intractable, costly, and/or noisy. This type of problem can be found in numerous real-world scenarios, including stochastic optimization and reinforcement learning, where each evaluation of a density

function may incur some computationally-expensive or even physical (real-world activity) cost, likely to give different results each time. The surrogate model does not incur this cost, but there are important trade-offs and considerations involved in the choice and design of such methodologies. We classify the different methodologies into three main classes and describe specific instances of algorithms under a unified notation. A modular scheme which encompasses the considered methods is also presented. A range of application scenarios is discussed, with special attention to the likelihood-free setting and reinforcement learning. Several numerical comparisons are also provided.

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.
This chapter examines the use and merits of hierarchical clustering techniques in the context of multi-asset multi-factor investing. In particular, it contrasts these techniques with several competing risk-based allocation paradigms, such as 1/N, minimum-variance, standard risk parity and diversified risk parity. The chapter introduces hierarchical risk parity (HRP) strategies based on the Pearson correlation coefficient and also introduces hierarchical clustering based on the lower tail dependence coefficient. The chapter provides an overview of traditional risk-based allocation strategies and outlines a framework to measure and manage portfolio diversification. It examines the performance of the introduced HRP strategies relative to the traditional alternatives. The chapter discusses Meucci's approach to managing diversification, which serves to construct a diversified risk parity strategy based on economic factors.

Loon, K. W., Graesser, L., and Cvitkovic, M. (2019). "SLM Lab: A Comprehensive Benchmark and Modular Software Framework for Reproducible Deep Reinforcement Learning." In: *arXiv e-Print*.
We introduce SLM Lab, a software framework for reproducible reinforcement learning (RL) research. SLM Lab implements a number of popular RL algorithms, provides synchronous and asynchronous parallel experiment execution, hyperparameter search, and result analysis. RL algorithms in SLM Lab are implemented in a modular way such that differences in algorithm performance can be confidently ascribed to differences between algorithms, not between implementations. In this work we present the design choices behind SLM Lab and use it to produce a comprehensive single-codebase RL algorithm benchmark. In addition, as a consequence of SLM Lab's modular design, we introduce and evaluate a discrete-action variant of the Soft Actor-Critic algorithm (Haarnoja et al., 2018) and a hybrid synchronous/asynchronous training method for RL agents.

Lopez de Prado, M. (2019). "A Data Science Solution to the Multiple-Testing Crisis in Financial Research." In: *The Journal of Financial Data Science* 1(1), pp. 99–110.
Most discoveries in empirical finance are false, as a consequence of selection bias under multiple testing. Although many researchers are aware of this problem, the solutions proposed in the literature tend to be complex and hard to implement. In this article, the author reduces the problem of selection bias in the context of investment strategy development to two sub-problems: determining the number of essentially independent trials and determining the variance across those trials. The author explains what data researchers need to report to allow others to evaluate the effect that multiple testing has had on reported performance. He applies his method to a real case of strategy development and estimates the probability that a discovered strategy is false.

Lopez de Prado, M. and Lewis, M. J. (2019). "Detection of false investment strategies using unsupervised learning methods." In: *Quantitative Finance* 19(9), pp. 1555–1565.
In this paper we address the problem of selection bias under multiple testing in the context of investment strategies. We introduce an unsupervised learning algorithm that determines the number of effectively uncorrelated trials carried out in the context of a discovery. This estimate is critical for computing the familywise false positive probability, and for filtering out false investment strategies.

Ma, Y. c. (, Wang, Z., and Fleiss, A. (2021). "Deep Q-Learning for Trading Cryptocurrency." In: *The Journal of Financial Data Science* 3(3), pp. 121–127.
This article sets forth a framework for deep reinforcement learning as applied to trading cryptocurrencies. Specifically, the authors adopt Q-Learning, which is a model-free reinforcement learning algorithm, to implement a deep neural network to approximate the best possible states and actions to take in the cryptocurrency market. Bitcoin, Ethereum, and Litecoin were selected as representatives to test the model. The Deep Q trading agent generated an average portfolio return of 65.98%, although it showed extreme volatility over the 2,000 runs. Despite the high volatility of deep reinforcement learning, the experiment demonstrates that it

has exceptionally high potential to be employed and provides a solid foundation on which to build further research.

Majid, A. Y., Saaybi, S., Rietbergen, T. van, Francois-Lavet, V., Prasad, R. V., and Verhoeven, C. (2021). "Deep Reinforcement Learning Versus Evolution Strategies: A Comparative Survey." In: *arXiv e-Print*.
Deep Reinforcement Learning (DRL) and Evolution Strategies (ESs) have surpassed human-level control in many sequential decision-making problems, yet many open challenges still exist. To get insights into the strengths and weaknesses of DRL versus ESs, an analysis of their respective capabilities and limitations is provided. After presenting their fundamental concepts and algorithms, a comparison is provided on key aspects such as scalability, exploration, adaptation to dynamic environments, and multi-agent learning. Then, the benefits of hybrid algorithms that combine concepts from DRL and ESs are highlighted. Finally, to have an indication about how they compare in real-world applications, a survey of the literature for the set of applications they support is provided.

Malavasi, M., Lozza, S. O., and Truck, S. (2021). "Second order of stochastic dominance efficiency vs mean variance efficiency." In: *European Journal of Operational Research* 290(3), pp. 1192–1206.
In this paper, we compare two of the main paradigms of portfolio theory: mean variance analysis and expected utility. In particular, we show empirically that mean variance efficient portfolios are typically sub-optimal for non satiable and risk averse investors. We illustrate that the second order stochastic dominance (SSD) efficient set is the solution of a multi-objective optimization problem. We further show that the market portfolio is not necessarily a solution to this optimization problem. We also conduct an empirical analysis, examining the ex ante and ex post performance of SSD and mean variance efficient portfolios, using a bootstrap approach. In an ex ante analysis, we compare empirical moments, the level of diversification and set distances of mean variance and SSD efficient sets. We also show that the global minimum variance (GMV) portfolio and the part of the mean variance efficient frontier (MVEF) composed of highly diversified portfolios is second order stochastically dominated. This result also provides a possible alternative explanation for the diversification puzzle. Conducting an ex post analysis, we construct second order stochastic dominating strategies that outperform the GMV portfolio in terms of wealth and various other performance measures, producing a positive ex post opportunity cost.

Martellini, L., Milhau, V., and Mulvey, J. (2020). "Securing Replacement Income with Goal-Based Retirement Investing Strategies." In: *The Journal of Retirement* 7(4), pp. 8–26.
Individuals preparing for retirement are currently left with an unsatisfactory choice between security with no flexibility with annuity products and flexibility without security with investment products such as balanced funds or target date funds. To get out of this impasse, the authors introduce a range of flexicure retirement goal-based investing strategies that offer both security and flexibility with respect to the objective of generating replacement income in decumulation. Recent advances in financial engineering and digital technologies make it possible to apply goal-based investing principles to a much broader population of investors than the few traditional clients who can afford customized mandates or private banking services, which suggests that these flexicure retirement solutions can be used as part of the solution to the global pension crisis.

Marzban, S., Delage, E., Li, J. Y., Desgagne-Bouchard, J., and Dussault, C. (2021). "WaveCorr: Correlation-savvy Deep Reinforcement Learning for Portfolio Management." In: *arXiv e-Print*.
The problem of portfolio management represents an important and challenging class of dynamic decision making problems, where rebalancing decisions need to be made over time with the consideration of many factors such as investors preferences, trading environments, and market conditions. In this paper, we present a new portfolio policy network architecture for deep reinforcement learning (DRL)that can exploit more effectively cross-asset dependency information and achieve better performance than state-of-the-art architectures. In particular, we introduce a new property, referred to as *asset permutation invariance*, for portfolio policy networks that exploit multi-asset time series data, and design the first portfolio policy network, named WaveCorr, that preserves this invariance property when treating asset correlation information. At the core of our design is an innovative permutation invariant correlation processing layer. An extensive set of experiments are conducted using data from both Canadian (TSX) and American stock markets (S&P 500), and WaveCorr consistently outperforms other architectures with an impressive 3%-25% absolute improvement in terms of average annual return, and up to more than 200% relative improvement in average Sharpe ratio. We also measured an improvement of a factor of up to 5 in the stability of performance under random choices of initial asset ordering and weights. The stability of the network has been found as particularly valuable by our industrial partner.

Moerland, T. M., Broekens, J., and Jonker, C. M. (2021). "Model-based Reinforcement Learning: A Survey." In: *arXiv e-Print*.
Sequential decision making, commonly formalized as Markov Decision Process (MDP) optimization, is a key challenge in artificial intelligence. Two key approaches to this problem are reinforcement learning (RL) and planning. This paper presents a survey of the integration of both fields, better known as model-based reinforcement learning. Model-based RL has two main steps. First, we systematically cover approaches to dynamics model learning, including challenges like dealing with stochasticity, uncertainty, partial observability, and temporal abstraction. Second, we present a systematic categorization of planning-learning integration, including aspects like: where to start planning, what budgets to allocate to planning and real data collection, how to plan, and how to integrate planning in the learning and acting loop. After these two section, we also discuss implicit model-based RL as an end-to-end alternative for model learning and planning, and we cover the potential benefits of model-based RL, like enhanced data efficiency, targeted exploration, and improved stability. The survey also draws connection to several related RL fields, like hierarchical RL and transfer. Altogether, the survey presents a broad conceptual overview of planning-learning combinations for MDP optimization.

Mohammed, S., Bealer, R., and Cohen, J. (2021). "Embracing advanced AI/ML to help investors achieve success: Vanguard Reinforcement Learning for Financial Goal Planning." In: *arXiv e-Print*.
In the world of advice and financial planning, there is seldom one right answer. While traditional algorithms have been successful in solving linear problems, its success often depends on choosing the right features from a dataset, which can be a challenge for nuanced financial planning scenarios. Reinforcement learning is a machine learning approach that can be employed with complex data sets where picking the right features can be nearly impossible. In this paper, we will explore the use of machine learning for financial forecasting, predicting economic indicators, and creating a savings strategy. Vanguard ML algorithm for goals-based financial planning is based on deep reinforcement learning that identifies optimal savings rates across multiple goals and sources of income to help clients achieve financial success. Vanguard learning algorithms are trained to identify market indicators and behaviors too complex to capture with formulas and rules, instead, it works to model the financial success trajectory of investors and their investment outcomes as a Markov decision process. We believe that reinforcement learning can be used to create value for advisors and end-investors, creating efficiency, more personalized plans, and data to enable customized solutions.

Mooney, T., Rapaka, R., and Vera, T. (2020). "Dynamic Regime Strategy for Stress Testing and Optimizing Institutional Investor Portfolios." In: *SSRN e-Print*.
Our work aims to develop a stand-alone trading system to construct portfolios that show the benefits of value and momentum style integration and presents the effectiveness of alternative integration methods for long-only absolute return funds, which seeks absolute returns that are not highly correlated to traditional assets such as stocks and bonds. Our approach uses the CRoss Industry Standard Process for Data Mining (CRISP-DM) model to guide the necessary steps, processes, and workflows for executing our project.

Mosavi, A., Ghamisi, P., Faghan, Y., Duan, P., ardabili, sina faizollahzadeh, Salwana, E., and Band, S. (2020). "Comprehensive Review of Deep Reinforcement Learning Methods and Applications in Economics." In: *SSRN e-Print*.
The popularity of deep reinforcement learning (DRL) methods in economics have been exponentially increased. DRL, through a wide range of capabilities from reinforcement learning (RL) to deep learning (DL), offers vast opportunities for handling sophisticated economics dynamic systems. DRL is characterized by scalability with the potential to be applied to high-dimensional problems in conjunction with noisy and nonlinear patterns of economic data. In this paper, we initially consider a brief review of DL, RL, and deep RL methods in diverse applications in economics, providing an in-depth insight into state of the art. Furthermore, the architecture of DRL applied to economic applications is investigated in order to highlight the complexity, robustness, accuracy, performance, computational tasks, risk constraints, and profitability. The survey results indicate that DRL can provide better performance and higher efficiency as compared to the traditional algorithms while facing real economic problems at the presence of risk parameters and the ever-increasing uncertainties.

Mulvey, J. M., Martellini, L., Hao, H., and Li, N. (2019). "A Factor- and Goal-Driven Model for Defined Benefit Pensions: Setting Realistic Benefits." In: *The Journal of Portfolio Management* 45 (3), pp. 165–177.
A factor and goal-driven framework for assessing asset allocation and contribution decisions within defined-benefit pension plans is developed in this article. A critical element is setting future benefits with reference

to the ability of the pension sponsors to support liabilities under reasonable investment expectations. The approach suggested by the authors combines a micro study of a representative cohort of individuals with an aggregation across a target population to estimate consistency between the micro and macro environments. A stochastic inflation risk factor affects both contribution and spending cash flows. This agent-based model suggested by the authors provides a more realistic framework than traditional approaches for setting pension benefits.

Muralidhar, A. (2020). "Asset Pricing, Asset Allocation and Risk-Adjusted Performance with Multiple Goals and Agency: The Goals and Risk-based Asset Pricing Model." In: *SSRN e-Print*.
Investment managers require a consistent asset pricing model, asset allocation recommendations and risk-adjusted performance measures (or the "three facets of investing") to be effective in managing portfolios. Incorporating three critical realities of investing into these models (i.e., that investors have many stochastic goals, seek to delegate to skillful agents, and maximize risk-adjusted returns as opposed to expected utility) provides recommendations on the three facets that are markedly different from the foundational papers of Modern Portfolio Theory (MPT). This is important as Goals-based Investing (GBI) and delegation are now the norm, and investors globally are not meeting their goals by adopting traditional MPT. The paper briefly surveys the literature on MPT, GBI, and agency before providing a normative Goals- and Risk-Based Asset Pricing Model (GRAPM) that includes these three realities of investing and articulates the three new facets. GRAPM exploits a simple idea that a relatively risk-free asset for one stochastic goal is a risky asset for another, and vice versa. These two assets, plus the traditional absolute risk-free rate of MPT, allow us to triangulate to establish returns for all other assets based on the return of any goal-replicating asset and multiple correlations (as opposed to a single relationship with the unobservable "market" portfolio). This approach creates a "pair-wise equilibrium" for all assets - very different from MPT - and also lends itself easily to a new asset pricing model with heterogeneous investors (i.e., each investor has a unique goal). GRAPM incorporates a "risk-aversion" parameter that is also easily observable, unlike MPT, and appears to explain why seemingly similar investors can have markedly different asset allocations or expected returns.

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*.
Hierarchical reinforcement learning has demonstrated significant success at solving difficult reinforcement learning (RL) tasks. Previous works have motivated the use of hierarchy by appealing to a number of intuitive benefits, including learning over temporally extended transitions, exploring over temporally extended periods, and training and exploring in a more semantically meaningful action space, among others. However, in fully observed, Markovian settings, it is not immediately clear why hierarchical RL should provide benefits over standard "shallow" RL architectures. In this work, we isolate and evaluate the claimed benefits of hierarchical RL on a suite of tasks encompassing locomotion, navigation, and manipulation. Surprisingly, we find that most of the observed benefits of hierarchy can be attributed to improved exploration, as opposed to easier policy learning or imposed hierarchical structures. Given this insight, we present exploration techniques inspired by hierarchy that achieve performance competitive with hierarchical RL while at the same time being much simpler to use and implement.

Naeem, M., Rizvi, S. T. H., and Coronato, A. (2020). "A Gentle Introduction to Reinforcement Learning and its Application in Different Fields." In: *IEEE Access* 8, pp. 209320–209344.
Due to the recent progress in Deep Neural Networks, Reinforcement Learning (RL) has become one of the most important and useful technology. It is a learning method where a software agent interacts with an unknown environment, selects actions, and progressively discovers the environment dynamics. RL has been effectively applied in many important areas of real life. This article intends to provide an in-depth introduction of the Markov Decision Process, RL and its algorithms. Moreover, we present a literature review of the application of RL to a variety of fields, including robotics and autonomous control, communication and networking, natural language processing, games and self-organized system, scheduling management and configuration of resources, and computer vision.

Nguyen, N. D., Nguyen, T. T., Nguyen, H., and Nahavandi, S. (2021). "Review, Analyze, and Design a Comprehensive Deep Reinforcement Learning Framework." In: *arXiv e-Print*.
Reinforcement learning (RL) has emerged as a standard approach for building an intelligent system, which involves multiple self-operated agents to collectively accomplish a designated task. More importantly, there has been a great attention to RL since the introduction of deep learning that essentially makes RL feasible to operate in high-dimensional environments. However, current research interests are diverted into different

directions, such as multi-agent and multi-objective learning, and human-machine interactions. Therefore, in this paper, we propose a comprehensive software architecture that not only plays a vital role in designing a connect-the-dots deep RL architecture but also provides a guideline to develop a realistic RL application in a short time span. By inheriting the proposed architecture, software managers can foresee any challenges when designing a deep RL-based system. As a result, they can expedite the design process and actively control every stage of software development, which is especially critical in agile development environments. For this reason, we designed a deep RL-based framework that strictly ensures flexibility, robustness, and scalability. Finally, to enforce generalization, the proposed architecture does not depend on a specific RL algorithm, a network configuration, the number of agents, or the type of agents.

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

Reinforcement learning is a machine learning approach concerned with solving dynamic optimization problems in an almost model-free way by maximizing a reward function in state and action spaces. This property makes it an exciting area of research for financial problems. Asset allocation, where the goal is to obtain the weights of the assets that maximize the rewards in a given state of the market considering risk and transaction costs, is a problem easily framed using a reinforcement learning framework. It is first a prediction problem for expected returns and covariance matrix and then an optimization problem for returns, risk, and market impact. Investors and financial researchers have been working with approaches like mean-variance optimization, minimum variance, risk parity, and equally weighted and several methods to make expected returns and covariance matrices' predictions more robust. This paper demonstrates the application of reinforcement learning to create a financial model-free solution to the asset allocation problem, learning to solve the problem using time series and deep neural networks. We demonstrate this on daily data for the top 24 stocks in the US equities universe with daily rebalancing. We use a deep reinforcement model on US stocks using different architectures. We use Long Short Term Memory networks, Convolutional Neural Networks, and Recurrent Neural Networks and compare them with more traditional portfolio management. The Deep Reinforcement Learning approach shows better results than traditional approaches using a simple reward function and only being given the time series of stocks. In Finance, no training to test error generalization results come guaranteed. We can say that the modeling framework can deal with time series prediction and asset allocation, including transaction costs.

Odermatt, L., Beqiraj, J., and Osterrieder, J. (2021). "Deep Reinforcement Learning for Finance and the Efficient Market Hypothesis." In: *SSRN e-Print*.

Is there an informational gain by training a Deep Reinforcement Learning agent for automated stock trading using other time series than the one to be traded? In this work, we implement a DRL algorithm in a solid framework within a model-free and actor-critic approach and learn it with 21 global Multi Assets to predict and trade on the S&P 500. The Efficient Market Hypothesis sets out that it is impossible to gather more information from the broader input. We demand to learn a DRL agent on this index with and without the additional information of these several Multi Assets to determine if the agent could capture invisible dependencies to end up with an informational gain and a better performance. The aim of this work is not to tune the hyperparameters of a DRL agent; several papers already exist on this subject. Nevertheless, we use a proven setup as model architecture. We take a Multi Layer Perceptron (short: MLP) as the neural network architecture with two hidden layers and 64 neurons each layer. The activation function used is the hyperbolic tangent. Further, Proximal Policy Optimization (short: PPO) is used as the policy for simple implementation and enabling a continuous state space. To deal with uncertainties of neural nets, we learn 100 agents for each scenario and compared both results. Neither the Sharpe ratios nor the cumulative returns are better in the more complex approach with the additional information of the Multi Assets, and even the single approach performed marginally better. However, we demonstrate that the complexly learned agent delivers less scattering over the 100 simulations in terms of the risk-adjusted returns, so there is an informational gain due to Multi Assets. A DRL agent learned with additional information delivers more robust results compared to the taken risk. We deliver valuable results for the further development of Deep Reinforcement Learning and provide a unique and resourceful approach.

Pardo, F. (2021). "Tonic: A Deep Reinforcement Learning Library for Fast Prototyping and Benchmarking." In: *arXiv e-Print*.

Deep reinforcement learning has been one of the fastest growing fields of machine learning over the past years and numerous libraries have been open sourced to support research. However, most codebases have a steep

learning curve or limited flexibility that do not satisfy a need for fast prototyping in fundamental research. This paper introduces Tonic, a Python library allowing researchers to quickly implement new ideas and measure their importance by providing: 1) general-purpose configurable modules 2) several baseline agents: A2C, TRPO, PPO, MPO, DDPG, D4PG, TD3 and SAC built with these modules 3) support for TensorFlow 2 and PyTorch 4) support for continuous-control environments from OpenAI Gym, DeepMind Control Suite and PyBullet 5) scripts to experiment in a reproducible way, plot results, and play with trained agents 6) a benchmark of the provided agents on 70 continuous-control tasks. Evaluation is performed in fair conditions with identical seeds, training and testing loops, while sharing general improvements such as non-terminal timeouts and observation normalization. Finally, to demonstrate how Tonic simplifies experimentation, a novel agent called TD4 is implemented and evaluated.

Parker, F. J. (2016a). "Goal-Based Portfolio Optimization." In: *The Journal of Wealth Management* 19(3), pp. 22–30.
The article presents a goal-based portfolio optimization approach that is truly native to the goal-based environment. It begins by redefining risk as the probability of failing to attain a specified goal and redefining reward as the excess wealth over and above what is required by the goal. It then presents an optimization procedure that seeks to minimize goal failure and maximize excess return. In preliminary tests, it finds that this goal-based procedure lowers the probability of failing to achieve a specified goal while delivering higher excess wealth than the procedures currently available.

Parker, F. J. (2016b). "Portfolio Selection in a Goal-Based Setting." In: *The Journal of Wealth Management* 19(2), pp. 41–46.
Using different portfolio ratios, we illustrate the deficiencies of using only modern portfolio theory (MPT) metrics and assumptions when selecting portfolios in a goal-based setting. Through the lenses of the ratios, we show how MPT can choose the wrong, albeit efficient, portfolio to accomplish a goal. These facts illustrate the need for goal-based practitioners to factor in other variables, such as time to a goal and maximum loss thresholds, when managing a portfolio to a goal-oriented mandate.

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.
Although the goals-based investment literature has grown, there remain two unsolved problems. First, there is no cohesive theory for the allocation of wealth across goals. If, for example, a client wants to retire in thirty years, send a child to university in eight years, and buy a vacation home in four, how she should divvy her wealth across those goals has been an open question. Restating the same problem: The vesting of shorter-dated goals carries a loss of achievement probability for longer-dated ones. How much probability loss is acceptable? Second, mean-variance portfolios yield lower probabilities of goal achievement than goals-based portfolios. I demonstrate use of the goals-based method. Parker (2020) introduced a cohesive goals-based allocation model that solves these problems. The approach, however, carries some practical challenges that are addressed in this discussion. Finally, I discuss some possible implications of the approach on the structure of firms, the regulatory environment, and the industry as a whole.

Parker, F. J. (2021a). "A Goals-Based Theory of Utility." In: *Journal of Behavioral Finance* 22(1), pp. 10–25.
Theoretical frameworks to date have prescribed how an investor should allocate wealth within mental accounts. However, there is no fully cohesive solution to prescribe how an investor should rationally allocate resources across mental accounts. It is the aim of this discussion to fill that theoretical gap. We present a framework which can be used to rationally allocate resources both within and across mental accounts or goals. We then compare and contrast this method with mean-variance optimization and behavioral portfolio theory, showing that both are stochastically dominated by the goals-based utility framework. In further analysis of empirical validity, we discuss the Samuelson Paradox, Friedman-Savage puzzle, and probability weighting functions.

Parker, F. J. (2021b). "Achieving Goals While Making an Impact: Balancing Financial Goals with Impact Investing." In: *The Journal of Impact and ESG Investing* 1(3), pp. 27–38.
For investors who wish to engage in impact investing and who have specific goals to achieve, there exists the potential for a trade-off. When impact investments yield lower returns than nonimpact portfolios, how much return should an investor be willing to give up to incorporate it? Using recent advances in goals-based utility theory, this article explores an answer to that question and offers practical and concrete advice for advisors to individual investors and fiduciaries of trusts. Using the goals-based framework, the author shows how an

investor's willingness to sacrifice return for an impact investing mandate changes in response to market and portfolio conditions.

Parker, F. J. (2021c). "Multi-Period Goals-Based Portfolio Optimization." In: *SSRN e-Print*.
Portfolio managers regularly have views on capital markets that span multiple periods. A portfolio manager, for example, may expect a recession with high probability over the coming period, followed by a recovery in the subsequent period. Currently, there are few methods for optimal portfolio allocation across these multiple periods that match how practitioners operate in the real world. Herein, we present a multi-period optimization method using a goals-based framework that allows practitioners to develop multi-period capital market expectations and optimize their portfolios across those periods.

Parker-Holder, J., Rajan, R., Song, X., Biedenkapp, A., Miao, Y., Eimer, T., Zhang, B., Nguyen, V., Calandra, R., Faust, A., Hutter, F., and Lindauer, M. (2022). "Automated Reinforcement Learning (AutoRL): A Survey and Open Problems." In: *arXiv e-Print*.
The combination of Reinforcement Learning (RL) with deep learning has led to a series of impressive feats, with many believing (deep) RL provides a path towards generally capable agents. However, the success of RL agents is often highly sensitive to design choices in the training process, which may require tedious and error-prone manual tuning. This makes it challenging to use RL for new problems, while also limits its full potential. In many other areas of machine learning, AutoML has shown it is possible to automate such design choices and has also yielded promising initial results when applied to RL. However, Automated Reinforcement Learning (AutoRL) involves not only standard applications of AutoML but also includes additional challenges unique to RL, that naturally produce a different set of methods. As such, AutoRL has been emerging as an important area of research in RL, providing promise in a variety of applications from RNA design to playing games such as Go. Given the diversity of methods and environments considered in RL, much of the research has been conducted in distinct subfields, ranging from meta-learning to evolution. In this survey we seek to unify the field of AutoRL, we provide a common taxonomy, discuss each area in detail and pose open problems which would be of interest to researchers going forward.

Pineda, L., Amos, B., Zhang, A., Lambert, N. O., and Calandra, R. (2021). "MBRL-Lib: A Modular Library for Model-based Reinforcement Learning." In: *arXiv e-Print*.
Model-based reinforcement learning is a compelling framework for data-efficient learning of agents that interact with the world. This family of algorithms has many subcomponents that need to be carefully selected and tuned. As a result the entry-bar for researchers to approach the field and to deploy it in real-world tasks can be daunting. In this paper, we present MBRL-Lib – a machine learning library for model-based reinforcement learning in continuous state-action spaces based on PyTorch. MBRL-Lib is designed as a platform for both researchers, to easily develop, debug and compare new algorithms, and non-expert user, to lower the entry-bar of deploying state-of-the-art algorithms. MBRL-Lib is open-source at `https://github.com/facebookresearch/mbrl-lib`.

Plaat, A. (2022). "Deep Reinforcement Learning." In: *arXiv e-Print*.
Deep reinforcement learning has gathered much attention recently. Impressive results were achieved in activities as diverse as autonomous driving, game playing, molecular recombination, and robotics. In all these fields, computer programs have taught themselves to solve difficult problems. They have learned to fly model helicopters and perform aerobatic manoeuvers such as loops and rolls. In some applications they have even become better than the best humans, such as in Atari, Go, poker and StarCraft. The way in which deep reinforcement learning explores complex environments reminds us of how children learn, by playfully trying out things, getting feedback, and trying again. The computer seems to truly possess aspects of human learning; this goes to the heart of the dream of artificial intelligence. The successes in research have not gone unnoticed by educators, and universities have started to offer courses on the subject. The aim of this book is to provide a comprehensive overview of the field of deep reinforcement learning. The book is written for graduate students of artificial intelligence, and for researchers and practitioners who wish to better understand deep reinforcement learning methods and their challenges. We assume an undergraduate-level of understanding of computer science and artificial intelligence; the programming language of this book is Python. We describe the foundations, the algorithms and the applications of deep reinforcement learning. We cover the established model-free and model-based methods that form the basis of the field. Developments go quickly, and we also cover advanced topics: deep multi-agent reinforcement learning, deep hierarchical reinforcement learning, and deep meta learning.

Plaat, A., Kosters, W., and Preuss, M. (2021). "High-Accuracy Model-Based Reinforcement Learning, a Survey." In: *arXiv e-Print*.
Deep reinforcement learning has shown remarkable success in the past few years. Highly complex sequential decision making problems from game playing and robotics have been solved with deep model-free methods. Unfortunately, the sample complexity of model-free methods is often high. To reduce the number of environment samples, model-based reinforcement learning creates an explicit model of the environment dynamics. Achieving high model accuracy is a challenge in high-dimensional problems. In recent years, a diverse landscape of model-based methods has been introduced to improve model accuracy, using methods such as uncertainty modeling, model-predictive control, latent models, and end-to-end learning and planning. Some of these methods succeed in achieving high accuracy at low sample complexity, most do so either in a robotics or in a games context. In this paper, we survey these methods; we explain in detail how they work and what their strengths and weaknesses are. We conclude with a research agenda for future work to make the methods more robust and more widely applicable to other applications.

Platanakis, E., Sutcliffe, C. M., and Ye, X. (2021). "Horses for Courses: Mean-Variance for Asset Allocation and 1/N for Stock Selection." In: *European Journal of Operational Research* 288(1), pp. 302–317.
For various organizational reasons, large investors typically split their portfolio decision into two stages - asset allocation and stock selection. We hypothesise that mean-variance models are superior to equal weighting for asset allocation, while the reverse applies for stock selection, as estimation errors are less of a problem for mean-variance models when used for asset allocation than for stock selection. We confirm this hypothesis in separate analyses of US and international equities using four different types of mean-variance model (Bayes-Stein, Black-Litterman, Bayesian diffuse prior and Markowitz), a range of parameter settings, and a simulation analysis calibrated to US data.

Pretorius, A., Tessera, K.-a., Smit, A. P., Formanek, C., Grimbly, S. J., Eloff, K., Danisa, S., Francis, L., Shock, J., Kamper, H., Brink, W., Engelbrecht, H., Laterre, A., and Beguir, K. (2021). "Mava: a research framework for distributed multi-agent reinforcement learning." In: *arXiv e-Print*.
Breakthrough advances in reinforcement learning (RL) research have led to a surge in the development and application of RL. To support the field and its rapid growth, several frameworks have emerged that aim to help the community more easily build effective and scalable agents. However, very few of these frameworks exclusively support multi-agent RL (MARL), an increasingly active field in itself, concerned with decentralised decision-making problems. In this work, we attempt to fill this gap by presenting Mava: a research framework specifically designed for building scalable MARL systems. Mava provides useful components, abstractions, utilities and tools for MARL and allows for simple scaling for multi-process system training and execution, while providing a high level of flexibility and composability. Mava is built on top of DeepMind's Acme, and therefore integrates with, and greatly benefits from, a wide range of already existing single-agent RL components made available in Acme. Several MARL baseline systems have already been implemented in Mava. These implementations serve as examples showcasing Mava's reusable features, such as interchangeable system architectures, communication and mixing modules. Furthermore, these implementations allow existing MARL algorithms to be easily reproduced and extended. We provide experimental results for these implementations on a wide range of multi-agent environments and highlight the benefits of distributed system training.

Pricope, T.-V. (2021). "Deep Reinforcement Learning in Quantitative Algorithmic Trading: A Review." In: *arXiv e-Print*.
Algorithmic stock trading has become a staple in today's financial market, the majority of trades being now fully automated. Deep Reinforcement Learning (DRL) agents proved to be to a force to be reckon with in many complex games like Chess and Go. We can look at the stock market historical price series and movements as a complex imperfect information environment in which we try to maximize return - profit and minimize risk. This paper reviews the progress made so far with deep reinforcement learning in the subdomain of AI in finance, more precisely, automated low-frequency quantitative stock trading. Many of the reviewed studies had only proof-of-concept ideals with experiments conducted in unrealistic settings and no real-time trading applications. For the majority of the works, despite all showing statistically significant improvements in performance compared to established baseline strategies, no decent profitability level was obtained. Furthermore, there is a lack of experimental testing in real-time, online trading platforms and a lack of meaningful comparisons between agents built on different types of DRL or human traders. We conclude

that DRL in stock trading has showed huge applicability potential rivalling professional traders under strong assumptions, but the research is still in the very early stages of development.

Prollochs, N. and Feuerriegel, S. (2019). "ReinforcementLearning: A Package to Perform Model-Free Reinforcement Learning in R." In: *The Journal of Open Source Software* 4(38), p. 1087.
Reinforcement learning refers to a group of methods from artificial intelligence where an agentperforms learning through trial and error (Sutton & Barto, 1998). It differs from supervisedlearning, since reinforcement learning requires no explicit labels; instead, the agent interactscontinuously with its environment. That is, the agent starts in a specific state and thenperforms an action, based on which it transitions to a new state and, depending on theoutcome, receives a reward. Different strategies (e.g. Q-learning) have been proposed tomaximize the overall reward, resulting in a so-called policy, which defines the best possibleaction in each state. As a main advantage, reinforcement learning is applicable to situationsin which the dynamics of the environment are unknown or too complex to evaluate (e.g. Mnihet al., 2015). However, there is currently no package available for performing reinforcement learning in R. As a remedy, we introduce the ReinforcementLearningR package, which allows an agent to learn the optimal behavior based on sampling experience consisting of states,actions and rewards (Prollochs and Feuerriegel, 2017). Based on such training examples, the package allows a reinforcement learning agent to learn an optimal policy that defines the best possible action in each state. Main features of ReinforcementLearning include, but arenot limited to: 1) Learning an optimal policy from a fixed set of a priori known transition samples 2) Predefined learning rules and action selection modes 3) A highly customizable framework for model-free reinforcement learning tasks.

Qian, H. and Yu, Y. (2021). "Derivative-Free Reinforcement Learning: A Review." In: *arXiv e-Print*.
Reinforcement learning is about learning agent models that make the best sequential decisions in unknown environments. In an unknown environment, the agent needs to explore the environment while exploiting the collected information, which usually forms a sophisticated problem to solve. Derivative-free optimization, meanwhile, is capable of solving sophisticated problems. It commonly uses a sampling-and-updating framework to iteratively improve the solution, where exploration and exploitation are also needed to be well balanced. Therefore, derivative-free optimization deals with a similar core issue as reinforcement learning, and has been introduced in reinforcement learning approaches, under the names of learning classifier systems and neuroevolution/evolutionary reinforcement learning. Although such methods have been developed for decades, recently, derivative-free reinforcement learning exhibits attracting increasing attention. However, recent survey on this topic is still lacking. In this article, we summarize methods of derivative-free reinforcement learning to date, and organize the methods in aspects including parameter updating, model selection, exploration, and parallel/distributed methods. Moreover, we discuss some current limitations and possible future directions, hoping that this article could bring more attentions to this topic and serve as a catalyst for developing novel and efficient approaches.

Radovanov, B. and Marcikic, A. (2014). "Testing The Performance Of The Investment Portfolio Using Block Bootstrap Method." In: *Economic Themes* 52(2).
The aim of this paper is to create a stable model of investment portfolio optimization through a high degree of diversification and reduction of sudden changes in the allocation with monitoring of the dynamics of the impact factor. In this sense, there is bootstrap application procedure, which, without an excessive number of constraints involved in the optimization process provides solutions based on uncertain information. Thus defined, the optimization method has been patented by Michaud (1999) entitled re-sampled efficiency. Accordingly, this paper offers a comparison of the performance block bootstrap optimization models and traditional Markowitz's model inside and outside of the sample by applying the most frequently traded stocks on the BSE. The results show a better performance out of the sample and the presence of a larger number of shares forming the portfolio through bootstrap methodology. However, only through the traditional optimization process could be attained optimum according to the required limits. Such effects can be observed by comparing the limits of efficiency obtained through these optimization models. However, optimization-based methods bootstrap finds its place in reducing errors of assessment resulting from the limited sample size.

Rafati, J. and Marcia, R. F. (2018). "Deep Reinforcement Learning via L-BFGS Optimization." In: *arXiv e-Print*.
Reinforcement Learning (RL) algorithms allow artificial agents to improve their action selections so as to increase rewarding experiences in their environments. Deep Reinforcement Learning algorithms require solving a nonconvex and nonlinear unconstrained optimization problem. Methods for solving the optimization problems in deep RL are restricted to the class of first-order algorithms, such as stochastic gradient descent

(SGD). The major drawback of the SGD methods is that they have the undesirable effect of not escaping saddle points and their performance can be seriously obstructed by ill-conditioning. Furthermore, SGD methods require exhaustive trial and error to fine-tune many learning parameters. Using second derivative information can result in improved convergence properties, but computing the Hessian matrix for large-scale problems is not practical. Quasi-Newton methods require only first-order gradient information, like SGD, but they can construct a low rank approximation of the Hessian matrix and result in superlinear convergence. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) approach is one of the most popular quasi-Newton methods that construct positive definite Hessian approximations. In this paper, we introduce an efficient optimization method, based on the limited memory BFGS quasi-Newton method using line search strategy – as an alternative to SGD methods. Our method bridges the disparity between first order methods and second order methods by continuing to use gradient information to calculate a low-rank Hessian approximations. We provide formal convergence analysis as well as empirical results on a subset of the classic ATARI 2600 games. Our results show a robust convergence with preferred generalization characteristics, as well as fast training time and no need for the experience replaying mechanism.

Raileanu, R., Goldstein, M., Yarats, D., Kostrikov, I., and Fergus, R. (2021). "Automatic Data Augmentation for Generalization in Deep Reinforcement Learning." In: *arXiv e-Print*.
Deep reinforcement learning (RL) agents often fail to generalize to unseen scenarios, even when they are trained on many instances of semantically similar environments. Data augmentation has recently been shown to improve the sample efficiency and generalization of RL agents. However, different tasks tend to benefit from different kinds of data augmentation. In this paper, we compare three approaches for automatically finding an appropriate augmentation. These are combined with two novel regularization terms for the policy and value function, required to make the use of data augmentation theoretically sound for certain actor-critic algorithms. We evaluate our methods on the Procgen benchmark which consists of 16 procedurally-generated environments and show that it improves test performance by 40% relative to standard RL algorithms. Our agent outperforms other baselines specifically designed to improve generalization in RL. In addition, we show that our agent learns policies and representations that are more robust to changes in the environment that do not affect the agent, such as the background. Our implementation is available at `https://github.com/rraileanu/auto-drac`.

Rebonato, R. (2019). "A financially justifiable and practically implementable approach to coherent stress testing." In: *Quantitative Finance* 19(5), pp. 827–842.
We present an approach to stress testing that is both practically implementable and solidly rooted in well-established financial theory. We present our results in a Bayesian-net context, but the approach can be extended to different settings. We show (i) how the consistency and continuity conditions are satisfied; (ii) how the result of a scenario can be consistently cascaded from a small number of macrofinancial variables to the constituents of a granular portfolio; and (iii) how an approximate but robust estimate of the likelihood of a given scenario can be estimated. This is particularly important for regulatory and capital-adequacy applications.

Ren, T., Zhang, T., Szepesvari, C., and Dai, B. (2021). "A Free Lunch from the Noise: Provable and Practical Exploration for Representation Learning." In: *arXiv e-Print*.
Representation learning lies at the heart of the empirical success of deep learning for dealing with the curse of dimensionality. However, the power of representation learning has not been fully exploited yet in reinforcement learning (RL), due to i), the trade-off between expressiveness and tractability; and ii), the coupling between exploration and representation learning. In this paper, we first reveal the fact that under some noise assumption in the stochastic control model, we can obtain the linear spectral feature of its corresponding Markov transition operator in closed-form for free. Based on this observation, we propose Spectral Dynamics Embedding (SPEDE), which breaks the trade-off and completes optimistic exploration for representation learning by exploiting the structure of the noise. We provide rigorous theoretical analysis of SPEDE, and demonstrate the practical superior performance over the existing state-of-the-art empirical algorithms on several benchmarks.

Roncalli, T. (2019). "How Machine Learning Can Improve Portfolio Allocation of Robo-Advisors." In: *SwissQuant Conference*.
solving portfolio optimization with machine learning algorithms.

Roncalli, T. (2021). "Advanced Course in Asset Management." In: *SSRN e-Print*.

These presentation slides have been written for the Advanced Course in Asset Management (theory and applications) given at the University of Paris-Saclay. They contain 15 tutorial exercises and 5 main lectures:

1) Portfolio Optimization
2) Risk Budgeting
3) Smart Beta, Factor Investing and Alternative Risk Premia
4) Green and Sustainable Finance, ESG Investing and Climate Risk
5) Machine Learning in Asset Management

The Table of contents is the following:
Part 1. Portfolio Optimization 1. Theory of portfolio optimization 1.a. The Markowitz framework 1.b. Capital asset pricing model (CAPM) 1.c. Portfolio optimization in the presence of a benchmark 1.d. Black-Litterman model 2. Practice of portfolio optimization 2.a. Covariance matrix 2.b. Expected returns 2.c. Regularization of optimized portfolios 2.d. Adding constraints 3. Tutorial exercises 3.a. Variations on the efficient frontier 3.b. Beta coefficient 3.c. Black-Litterman model
Part 2. Risk Budgeting 1. The ERC portfolio 1.a. Definition 1.b. Special cases 1.c. Properties 1.d. Numerical solution 2. Extensions to risk budgeting portfolios 2.a. Definition of RB portfolios 2.b. Properties of RB portfolios 2.c. Diversification measures 2.d. Using risk factors instead of assets 3. Risk budgeting, risk premia and the risk parity strategy 3.a. Diversified funds 3.b. Risk premium 3.c. Risk parity strategies 3.d. Performance budgeting portfolios 4. Tutorial exercises 4.a. Variation on the ERC portfolio 4.b. Weight concentration of a portfolio 4.c. The optimization problem of the ERC portfolio 4.d. Risk parity funds
Part 3. Smart Beta, Factor Investing and Alternative Risk Premia 1. Risk-based indexation 1.a. Capitalization-weighted indexation 1.b. Risk-based portfolios 1.c. Comparison of the four risk-based portfolios 1.d. The case of bonds 2. Factor investing 2.a. Factor investing in equities 2.b. How many risk factors? 2.c. Construction of risk factors 2.d. Risk factors in other asset classes 3. Alternative risk premia 3.a. Definition 3.b. Carry, value, momentum and liquidity 3.c. Portfolio allocation with ARP 4. Tutorial exercises 4.a. Equally-weighted portfolio 4.b. Most diversified portfolio 4.c. Computation of risk-based portfolios 4.d. Building a carry trade exposure
Part 4. Green and Sustainable Finance, ESG Investing and Climate Risk 1. ESG investing 1.a. Introduction to sustainable finance 1.b. ESG scoring 1.c. Performance in the stock market 1.d. Performance in the corporate bond market 2. Climate risk 2.a. Introduction to climate risk 2.b. Climate risk modeling 2.c. Regulation of climate risk 2.d. Portfolio management with climate risk 3. Sustainable financing products 3.a. SRI Investment funds 3.b. Green bonds 3.c. Social bonds 3.d. Other sustainability-linked strategies 4. Impact investing 4.a. Definition 4.b. Sustainable development goals (SDG) 4.c. Voting policy, shareholder activism and engagement 4.d. The challenge of reporting 5. Tutorial exercises 5.a. Probability distribution of an ESG score 5.b. Enhanced ESG score and tracking error control
Part 5. Machine Learning in Asset Management 1. Portfolio optimization 1.a. Standard optimization algorithms 1.b. Machine learning optimization algorithms 1.c. Application to portfolio allocation 2. Pattern learning and self-automated strategies 3. Market generators 4. Tutorial exercises 4.a. Portfolio optimization with CCD and ADMM algorithms 4.b. Regularized portfolio optimization.

Sato, Y. (2019). "Model-Free Reinforcement Learning for Financial Portfolios: A Brief Survey." In: *arXiv e-Print*. Financial portfolio management is one of the problems that are most frequently encountered in the investment industry. Nevertheless, it is not widely recognized that both Kelly Criterion and Risk Parity collapse into Mean Variance under some conditions, which implies that a universal solution to the portfolio optimization problem could potentially exist. In fact, the process of sequential computation of optimal component weights that maximize the portfolio's expected return subject to a certain risk budget can be reformulated as a discrete-time Markov Decision Process (MDP) and hence as a stochastic optimal control, where the system being controlled is a portfolio consisting of multiple investment components, and the control is its component weights. Consequently, the problem could be solved using model-free Reinforcement Learning (RL) without knowing specific component dynamics. By examining existing methods of both value-based and policy-based model-free RL for the portfolio optimization problem, we identify some of the key unresolved questions and difficulties facing today's portfolio managers of applying model-free RL to their investment portfolios.

Schumann, E. (2019). "Backtesting." In: *SSRN e-Print*.

We discuss the backtesting of investment and trading strategies. We start with the challenges and pitfalls: overfitting, data preparation, and the effects of randomness. Then, we introduce and describe R software for backtesting. We demonstrate how to use the software for univariate and multivariate strategies (i.e. portfolio strategies) for two equity data sets. Specifically, we discuss the implementation and testing of momentum and portfolio optimization models. Throughout, we stress the analysis of sensitivity and robustness checks. Since such analyses require to run many backtests, we also discuss how backtests can be run in parallel.

Schwarzer, M., Rajkumar, N., Noukhovitch, M., Anand, A., Charlin, L., Hjelm, D., Bachman, P., and Courville, A. (2021). "Pretraining Representations for Data-Efficient Reinforcement Learning." In: *arXiv e-Print*.
Data efficiency is a key challenge for deep reinforcement learning. We address this problem by using unlabeled data to pretrain an encoder which is then finetuned on a small amount of task-specific data. To encourage learning representations which capture diverse aspects of the underlying MDP, we employ a combination of latent dynamics modelling and unsupervised goal-conditioned RL. When limited to 100k steps of interaction on Atari games (equivalent to two hours of human experience), our approach significantly surpasses prior work combining offline representation pretraining with task-specific finetuning, and compares favourably with other pretraining methods that require orders of magnitude more data. Our approach shows particular promise when combined with larger models as well as more diverse, task-aligned observational data – approaching human-level performance and data-efficiency on Atari in our best setting. We provide code associated with this work at `https://github.com/mila-iqia/SGI`.

Seymour, A., Flint, E. J., and Chikurunhe, F. (2018). "Dynamic portfolio management strategies: A framework for historical analysis." In: *SSRN e-Print*.
The performance of dynamic trading and investment strategies can be difficult to predict. Although not without its problems, analysis of the historical performance of a strategy can provide valuable insight into its general risk and return properties. Furthermore, historical analysis allows one to compare variations of a strategy and examine the impact of various parameter choices and implementation rules. Dynamic strategy applications in three areas are considered, namely derivatives, asset allocation and equity factor portfolios. Firstly, the analysis of a strategy involving single-stock derivatives is examined in which call options on certain constituents of an index portfolio are sold as an alternative method of under-weighting the underlying. Secondly, the historical performance of an optimization-based asset allocation strategy is considered. The assumed aim of the strategy is to outperform a benchmark of CPI 5 via dynamic trading in a portfolio of domestic equities, bonds, property and cash, as well as international equities and bonds. Finally, the effects of portfolio construction on factor performance are studied via an historical analysis in which portfolios corresponding to a selection of fundamental factors are managed according to a range of weighting schemes, rebalance frequencies and portfolio sizes.

Shalett, L. (2015). *An Outcomes-Oriented Approach to Alternatives*. Tech. rep. Morgan Stanley Wealth Management.
Transformational forces are colliding in a way that necessitates a fresh approach to asset allocation guidance for alternative asset classes and strategies: the proliferation of lower-cost alternative investment formats; the normalization of interest rates; and the need to reintroduce alternatives to Financial Advisors and clients, many of whom in the past have been disillusioned by unfulfilled expectations, high fees, tax complexity and liquidity. In this new outcomes-based approach, we have a navigation framework that is intuitive and tests for suitability through alignment with basic portfolio goals. We also posit performance parameters that allow us to compare the trade-offs between alternative mutual funds/ ETFs and private offerings and suggest benchmarks that provide clients with a way to measure success.

Shi, W., Song, S., Wang, Z., and Huang, G. (2020). "Self-Supervised Discovering of Causal Features: Towards Interpretable Reinforcement Learning." In: *arXiv e-Print*.
Deep reinforcement learning (RL) has recently led to many breakthroughs on a range of complex control tasks. However, the agent's decision-making process is generally not transparent. The lack of interpretability hinders the applicability of RL in safety-critical scenarios. In this paper, we propose a self-supervised interpretable framework, which employs a self-supervised interpretable network (SSINet) to discover and locate fine-grained causal features that constitute most evidence for the agent's decisions. We verify and evaluate our method on several Atari 2600 games as well as Duckietown. The results show that our method renders causal explanations and empirical evidences about how the agent makes decisions and why the agent performs well or badly. Moreover, our method is a flexible explanatory module that can be applied to most vision-based RL agents. Overall, our method provides valuable insight into interpretable vision-based RL.

Soleymani, F. and Paquet, E. (2021). "Deep Graph Convolutional Reinforcement Learning for Financial Portfolio Management - DeepPocket." In: *Expert Systems with Applications*.
Portfolio management aims at maximizing the return on investment while minimizing risk by continuously reallocating the assets forming the portfolio. These assets are not independent but correlated during a short time period. A graph convolutional reinforcement learning framework called DeepPocket is proposed whose objective is to exploit the time-varying interrelations between financial instruments. These interrelations are represented by a graph whose nodes correspond to the financial instruments while the edges correspond to a pair-wise correlation function in between assets. DeepPocket consists of a restricted, stacked autoencoder for feature extraction, a convolutional network to collect underlying local information shared among financial instruments and an actor-critic reinforcement learning agent. The actor-critic structure contains two convolutional networks in which the actor learns and enforces an investment policy which is, in turn, evaluated by the critic in order to determine the best course of action by constantly reallocating the various portfolio assets to optimize the expected return on investment. The agent is initially trained offline with online stochastic batching on historical data. As new data become available, it is trained online with a passive concept drift approach to handle unexpected changes in their distributions. DeepPocket is evaluated against five real-life datasets over three distinct investment periods, including during the Covid-19 crisis, and clearly outperformed market indexes.

Staley, E. W., Rivera, C. G., and Llorens, A. J. (2021). "The AI Arena: A Framework for Distributed Multi-Agent Reinforcement Learning." In: *arXiv e-Print*.
Advances in reinforcement learning (RL) have resulted in recent breakthroughs in the application of artificial intelligence (AI) across many different domains. An emerging landscape of development environments is making powerful RL techniques more accessible for a growing community of researchers. However, most existing frameworks do not directly address the problem of learning in complex operating environments, such as dense urban settings or defense-related scenarios, that incorporate distributed, heterogeneous teams of agents. To help enable AI research for this important class of applications, we introduce the AI Arena: a scalable framework with flexible abstractions for distributed multi-agent reinforcement learning. The AI Arena extends the OpenAI Gym interface to allow greater flexibility in learning control policies across multiple agents with heterogeneous learning strategies and localized views of the environment. To illustrate the utility of our framework, we present experimental results that demonstrate performance gains due to a distributed multi-agent learning approach over commonly-used RL techniques in several different learning environments.

Stapelberg, B. and Malan, K. M. (2021). "A survey of benchmarking frameworks for reinforcement learning." In: *arXiv e-Print*.
Reinforcement learning has recently experienced increased prominence in the machine learning community. There are many approaches to solving reinforcement learning problems with new techniques developed constantly. When solving problems using reinforcement learning, there are various difficult challenges to overcome. To ensure progress in the field, benchmarks are important for testing new algorithms and comparing with other approaches. The reproducibility of results for fair comparison is therefore vital in ensuring that improvements are accurately judged. This paper provides an overview of different contributions to reinforcement learning benchmarking and discusses how they can assist researchers to address the challenges facing reinforcement learning. The contributions discussed are the most used and recent in the literature. The paper discusses the contributions in terms of implementation, tasks and provided algorithm implementations with benchmarks. The survey aims to bring attention to the wide range of reinforcement learning benchmarking tasks available and to encourage research to take place in a standardised manner. Additionally, this survey acts as an overview for researchers not familiar with the different tasks that can be used to develop and test new reinforcement learning algorithms.

Stooke, A. and Abbeel, P. (2018). "Accelerated Methods for Deep Reinforcement Learning." In: *arXiv e-Print*.
Deep reinforcement learning (RL) has achieved many recent successes, yet experiment turn-around time remains a key bottleneck in research and in practice. We investigate how to optimize existing deep RL algorithms for modern computers, specifically for a combination of CPUs and GPUs. We confirm that both policy gradient and Q-value learning algorithms can be adapted to learn using many parallel simulator instances. We further find it possible to train using batch sizes considerably larger than are standard, without negatively affecting sample complexity or final performance. We leverage these facts to build a unified framework for parallelization that dramatically hastens experiments in both classes of algorithm. All neural network computations use GPUs, accelerating both data collection and training. Our results include using an entire

NVIDIA DGX-1 to learn successful strategies in Atari games in single-digit minutes, using both synchronous and asynchronous algorithms.

Stooke, A. and Abbeel, P. (2019). "rlpyt: A Research Code Base for Deep Reinforcement Learning in PyTorch." In: *arXiv e-Print*.
Since the recent advent of deep reinforcement learning for game play and simulated robotic control, a multitude of new algorithms have flourished. Most are model-free algorithms which can be categorized into three families: deep Q-learning, policy gradients, and Q-value policy gradients. These have developed along separate lines of research, such that few, if any, code bases incorporate all three kinds. Yet these algorithms share a great depth of common deep reinforcement learning machinery. We are pleased to share rlpyt, which implements all three algorithm families on top of a shared, optimized infrastructure, in a single repository. It contains modular implementations of many common deep RL algorithms in Python using PyTorch, a leading deep learning library. rlpyt is designed as a high-throughput code base for small- to medium-scale research in deep RL. This white paper summarizes its features, algorithms implemented, and relation to prior work, and concludes with detailed implementation and usage notes. rlpyt is available at `https://github.com/astooke/rlpyt`.

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.
The authors investigate the biases in the backtested performance of "alternative beta"strategies using a unique sample of 215 trading strategies developed and promoted by global investment banks. Their results lend support to the cautions in the recent literature regarding backtest overfitting and lack of robustness in trading strategy performance during the "live"period (out of sample). The authors report a median 73 percent deterioration in Sharpe ratios between backtested and live performance periods for the strategies, and they establish a link between performance deterioration and strategy complexity, with the realized reduction in live versus backtested Sharpe ratios of the most complex strategies exceeding those of the simplest ones by over 30 percentage points. The robustness of strategy exposure to risk factors varies between asset classes and strategies; it appears reasonable in equity volatility and FX carry strategies but quite weak in the equity value strategy in particular.

Sun, H., Zhong, J., Ma, Y., Han, Z., and He, K. (2021a). "TimeTraveler: Reinforcement Learning for Temporal Knowledge Graph Forecasting." In: *arXiv e-Print*.
Temporal knowledge graph (TKG) reasoning is a crucial task that has gained increasing research interest in recent years. Most existing methods focus on reasoning at past timestamps to complete the missing facts, and there are only a few works of reasoning on known TKGs to forecast future facts. Compared with the completion task, the forecasting task is more difficult that faces two main challenges: (1) how to effectively model the time information to handle future timestamps? (2) how to make inductive inference to handle previously unseen entities that emerge over time? To address these challenges, we propose the first reinforcement learning method for forecasting. Specifically, the agent travels on historical knowledge graph snapshots to search for the answer. Our method defines a relative time encoding function to capture the timespan information, and we design a novel time-shaped reward based on Dirichlet distribution to guide the model learning. Furthermore, we propose a novel representation method for unseen entities to improve the inductive inference ability of the model. We evaluate our method for this link prediction task at future timestamps. Extensive experiments on four benchmark datasets demonstrate substantial performance improvement meanwhile with higher explainability, less calculation, and fewer parameters when compared with existing state-of-the-art methods.

Sun, S., Wang, R., and An, B. (2021b). "Reinforcement Learning for Quantitative Trading." In: *arXiv e-Print*.
Quantitative trading (QT), which refers to the usage of mathematical models and data-driven techniques in analyzing the financial market, has been a popular topic in both academia and financial industry since 1970s. In the last decade, reinforcement learning (RL) has garnered significant interest in many domains such as robotics and video games, owing to its outstanding ability on solving complex sequential decision making problems. RL's impact is pervasive, recently demonstrating its ability to conquer many challenging QT tasks. It is a flourishing research direction to explore RL techniques' potential on QT tasks. This paper aims at providing a comprehensive survey of research efforts on RL-based methods for QT tasks. More concretely, we devise a taxonomy of RL-based QT models, along with a comprehensive summary of the state of the art. Finally, we discuss current challenges and propose future research directions in this exciting field.

Suri, K., Shi, X. Q., Plataniotis, K., and Lawryshyn, Y. (2021). "TradeR: Practical Deep Hierarchical Reinforcement Learning for Trade Execution." In: *arXiv e-Print*.

Advances in Reinforcement Learning (RL) span a wide variety of applications which motivate development in this area. While application tasks serve as suitable benchmarks for real world problems, RL is seldomly used in practical scenarios consisting of abrupt dynamics. This allows one to rethink the problem setup in light of practical challenges. We present Trade Execution using Reinforcement Learning (TradeR) which aims to address two such practical challenges of catastrophy and surprise minimization by formulating trading as a real-world hierarchical RL problem. Through this lens, TradeR makes use of hierarchical RL to execute trade bids on high frequency real market experiences comprising of abrupt price variations during the 2019 fiscal year COVID19 stock market crash. The framework utilizes an energy-based scheme in conjunction with surprise value function for estimating and minimizing surprise. In a large-scale study of 35 stock symbols from the S&P500 index, TradeR demonstrates robustness to abrupt price changes and catastrophic losses while maintaining profitable outcomes. We hope that our work serves as a motivating example for application of RL to practical problems.

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

Reinforcement learning, one of the most active research areas in artificial intelligence, is a computational approach to learning whereby an agent tries to maximize the total amount of reward it receives while interacting with a complex, uncertain environment. In Reinforcement Learning, Richard Sutton and Andrew Barto provide a clear and simple account of the field's key ideas and algorithms. This second edition has been significantly expanded and updated, presenting new topics and updating coverage of other topics. Like the first edition, this second edition focuses on core online learning algorithms, with the more mathematical material set off in shaded boxes. Part I covers as much of reinforcement learning as possible without going beyond the tabular case for which exact solutions can be found. Many algorithms presented in this part are new to the second edition, including UCB, Expected Sarsa, and Double Learning. Part II extends these ideas to function approximation, with new sections on such topics as artificial neural networks and the Fourier basis, and offers expanded treatment of off-policy learning and policy-gradient methods. Part III has new chapters on reinforcement learning's relationships to psychology and neuroscience, as well as an updated case-studies chapter including AlphaGo and AlphaGo Zero, Atari game playing, and IBM Watson's wagering strategy. The final chapter discusses the future societal impacts of reinforcement learning.

Taljaard, B. H. and Maré, E. (2021). "Why has the equal weight portfolio underperformed and what can we do about it?" In: *Quantitative Finance* 21(11), pp. 1855–1868.

It is widely noted that market capitalisation weighted portfolios are inefficient and underperform an equal weighted portfolio over the long-term. However, at least since 2016, an equal weighted portfolio of stocks in the S&P500 has significantly underperformed the market capitalisation weighted portfolio. In this paper, we analyse this underperformance using stochastic portfolio theory. We show that the equal weighted portfolio does appear to outperform the market capitalisation weighted portfolio over the long-term but with periods of significant short-term underperformance. In addition, we find that concentration in the market capitalisation weighted portfolio has increased in recent years and has contributed to the recent underperformance together with a significantly lower level of diversification benefits. Furthermore, we highlight an approach to improve the performance of a portfolio by dynamically selecting a market cap or an equal weighting using a rudimentary linear regression model.

Tarbouriech, J., Domingues, O. D., Menard, P., Pirotta, M., Valko, M., and Lazaric, A. (2021). "Adaptive Multi-Goal Exploration." In: *arXiv e-Print*.

We introduce a generic strategy for provably efficient multi-goal exploration. It relies on AdaGoal, a novel goal selection scheme that is based on a simple constrained optimization problem, which adaptively targets goal states that are neither too difficult nor too easy to reach according to the agent's current knowledge. We show how AdaGoal can be used to tackle the objective of learning an ϵ-optimal goal-conditioned policy for all the goal states that are reachable within L steps in expectation from a reference state s_0 in a reward-free Markov decision process. In the tabular case with S states and A actions, our algorithm requires $\tilde{O}(L^3SA\epsilon^{-2})$ exploration steps, which is nearly minimax optimal. We also readily instantiate AdaGoal in linear mixture Markov decision processes, which yields the first goal-oriented PAC guarantee with linear function approximation. Beyond its strong theoretical guarantees, AdaGoal is anchored in the high-level algorithmic structure of existing methods for goal-conditioned deep reinforcement learning.

Tayali, S. T. (2020). "A novel backtesting methodology for clustering in mean–variance portfolio optimization." In: *Knowledge-Based Systems* 209, p. 106454.

The decisions of asset selection and allocation lie at the heart of financial portfolio management. For these challenging tasks, the mathematical programming model of the mean-variance optimization problem proposes to use the concept of diversification. The novel methodology in this article is a representation of the accumulated knowledge of this model from the modern portfolio theory. It is a practical application for portfolio managers to help synthesize the available historical data and to infer rational decisions. The state-of-the-art backtesting methodology integrates the unsupervised machine learning method of clustering analysis into the mean-variance portfolio optimization model. The test results from the proposed novel methodology show that clustering with Euclidean distance measures outperform the results of the benchmark and other specified clustering methods for different datasets, backtesting periods, and temporal scales of major stock indices.

Terry, J. K., Black, B., Grammel, N., Jayakumar, M., Hari, A., Sullivan, R., Santos, L., Perez, R., Horsch, C., Dieffendahl, C., Williams, N. L., Lokesh, Y., and Ravi, P. (2021). "PettingZoo: Gym for Multi-Agent Reinforcement Learning." In: *arXiv e-Print*.

This paper introduces the PettingZoo library and the accompanying Agent Environment Cycle ("AEC") games model. PettingZoo is a library of diverse sets of multi-agent environments with a universal, elegant Python API. PettingZoo was developed with the goal of accelerating research in Multi-Agent Reinforcement Learning ("MARL"), by making work more interchangeable, accessible and reproducible akin to what OpenAI's Gym library did for single-agent reinforcement learning. PettingZoo's API, while inheriting many features of Gym, is unique amongst MARL APIs in that it's based around the novel AEC games model. We argue, in part through case studies on major problems in popular MARL environments, that the popular game models are poor conceptual models of the games commonly used with MARL, that they promote severe bugs that are hard to detect, and that the AEC games model addresses these problems.

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

Pascal Traccucci, Luc Dumontier, Guillaume Garchery and Benjamin Jacot present an extended reverse stress test (ERST) triptych approach with three variables: level of plausibility, level of loss and scenario. Any two of these variables can be derived, provided the third is given as input. A new version of the Levenberg-Marquardt optimisation algorithm is introduced to derive the ERST in certain complex cases.

Ungari, S. and Benhamou, E. (2021). "Deep Reinforcement Learning for Portfolio Allocation." In: *Risk Magazine Global Quant Network*.

In 2013, a paper by Google DeepMind kicked off an explosion in Deep Reinforcement Learning (DRL), for games. In this talk, we show that DRL can also be applied to portfolio allocation given various tricks and adaptation specific to non stationary data in finance. We present in particular how to Boost DRL.

Valentine, K. D., Buchanan, E. M., Scofield, J. E., and Beauchamp, M. T. (2019). "Beyond p values: utilizing multiple methods to evaluate evidence." In: *Behaviormetrika* 46(1), pp. 121–144.

Null hypothesis significance testing is cited as a threat to validity and reproducibility. While many individuals suggest that we focus on altering the p value at which we deem an effect significant, we believe this suggestion is short-sighted. Alternative procedures (i.e., Bayesian analyses and observation-oriented modeling: OOM) can be more powerful and meaningful to our discipline. However, these methodologies are less frequently utilized and are rarely discussed in combination with NHST. Herein, we discuss three methodologies (NHST, Bayesian Model comparison, and OOM), then compare the possible interpretations of three analyses (ANOVA, Bayes Factor, and an Ordinal Pattern Analysis) in various data environments using a frequentist simulation study. We found that changing significance thresholds had little effect on conclusions. Furthermore, we suggest that evaluating multiple estimates as evidence of an effect allows for more robust and nuanced interpretations of results and implies the need to redefine evidentiary value and reporting practices. Recent events in psychological science have prompted concerns within the discipline regarding research practices and ultimately, the validity and reproducibility of published reports (Etz and Vandekerckhove 2016; Lindsay 2015, Open Science Collaboration 2015; van Elk et al. 2015). One often discussed matter is over-reliance, abuse, and potential hacking of p values produced by frequentist null hypothesis significance testing (NHST), as well as misinterpretations of NHST results (Gigerenzer 2004; Ioannidis 2005; Simmons et al. 2011). We agree with these concerns and believe that many before us have voiced sound, generally accepted opinions on potential remedies, such as an increased focus on effect sizes (Cumming 2008; Lakens 2013; Maxwell et al. 2015; Nosek et al. 2012). However, other suggestions have been met with less enthusiasm, including an article by Benjamin et al. (2018) advocating that researchers should begin thinking only of p values less than .005 as "statistically significant", thus changing alpha levels to control Type I error rates. Alternatively,

Pericchi and Pereira (2016) promote the use of fluctuating alpha levels as a function of sample size to assist with these errors. Trafimow et al. (2018) critiques this suggestion to broadly lower the alpha level to .005 and suggested that findings should be weighted on the basis of evidence accumulation from multiple studies. We argue that alpha should not be the sole focus of our attention, but rather, we should wonder if a p value should be utilized at all, and, if so, what that p value can tell us in relation with other indicators. While NHST and p values may have merit, researchers have a wealth of other statistical tools available to them. We believe that improvements may be made to the sciences as a whole when individuals become aware of these tools and how these methods may be used, either alone or in combination, to strengthen understanding of data and conclusions. These sentiments have been shared by the American Statistical Association who recently held a conference focusing on going beyond NHST, expanding their previous stance on p values (Wasserstein and Lazar 2016). Therefore, the main goal of this project was to show researchers how two alternative paradigms compare to NHST in terms of their methodological design, statistical interpretations, and comparative robustness. Herein, we will discuss the following methodologies: NHST, Bayes factor comparisons, and observation-oriented modeling. To compare their methodological designs, we first provide historical backgrounds, procedural steps, and limitations for each paradigm. We then simulated data using a three timepoint repeated measures design with a Likert-type scale as the outcome variable to be able to compare the statistical interpretations and comparative robustness. By simulating possible data sets and analyzing them with each of the three paradigms, we will be able to discuss the conclusions these three methods reach given the same data and to compare how often these methodologies agree within different data environments (i.e., given varying sample sizes and effect sizes). Beyond simply comparing methodologies, we also sought to identify how changing the alpha criteria within the NHST framework may alter conclusions. Although previous work has already compared Frequentist NHST to Bayesian approaches (Goodman 1999; Rouder et al. 2012; Wetzels et al. 2011), this manuscript adds a novel contribution: observation-oriented modeling. By introducing social scientists to observation-oriented modeling (OOM), a relatively new paradigm that is readily interpretable, we will show both how useful this paradigm can be in these contexts, and how it compares to two well-known methods. We hope that by discussing these methodologies in terms of a simple statistical analysis researchers will be able to easily compare and contrast methodologies.

Vincent, K., Hsu, Y.-C., and Lin, H.-W. (2018). "Analyzing the Performance of Multifactor Investment Strategies under a Multiple Testing Framework." In: *The Journal of Portfolio Management* 44(4), pp. 113–126.
Evaluating portfolios based on numerous combinations of factors using the individual backtesting method could suffer from serious data mining bias and lead to spurious significant findings. Accordingly, the authors employ a multiple hypothesis testing method to examine the multifactor portfolio performance. Their empirical results show that even after they adjust for the multiple comparisons bias, stock-picking strategies with certain combined firm characteristics could generate significantly better liquidity risk-adjusted returns. In addition, the outperforming multifactor strategies that the authors report are robust to alternative definitions of factors. However, they observe that the number of significantly profitable multifactor portfolios has decreased substantially in the era of increased liquidity and trading activity in the U.S. stock market.

Vovk, V. and Wang, R. (2020). "True and false discoveries with e-values." In: *arXiv e-Print*.
The topic of this paper is multiple hypothesis testing based on e-values, which are Bayes factors stripped of their Bayesian content. Using e-values instead of p-values, which are standard in this area, leads to simple and efficient procedures that control the number of false discoveries under arbitrary dependence of the base e-values. We prove an optimality result for our main procedure and demonstrate advantages of our methods over standard methods using simulated and real-world datasets.

Vovk, V. and Wang, R. (2021). "E-values: Calibration, combination, and applications." In: *Annals of Statistics* 49(3), pp. 1736–1753.
Multiple testing of a single hypothesis and testing multiple hypotheses are usually done in terms of p-values. In this paper we replace p-values with their natural competitor, e-values, which are closely related to betting, Bayes factors, and likelihood ratios. We demonstrate that e-values are often mathematically more tractable; in particular, in multiple testing of a single hypothesis, e-values can be merged simply by averaging them. This allows us to develop efficient procedures using e-values for testing multiple hypotheses.

Wang, H. and Yu, S. (2021). "Robo-Advising: Enhancing Investment with Inverse Optimization and Deep Reinforcement Learning." In: *arXiv e-Print*.
Machine Learning (ML) has been embraced as a powerful tool by the financial industry, with notable applications spreading in various domains including investment management. In this work, we propose a full-cycle

data-driven investment robo-advising framework, consisting of two ML agents. The first agent, an inverse portfolio optimization agent, infers an investor's risk preference and expected return directly from historical allocation data using online inverse optimization. The second agent, a deep reinforcement learning (RL) agent, aggregates the inferred sequence of expected returns to formulate a new multi-period mean-variance portfolio optimization problem that can be solved using deep RL approaches. The proposed investment pipeline is applied on real market data from April 1, 2016 to February 1, 2021 and has shown to consistently outperform the S&P 500 benchmark portfolio that represents the aggregate market optimal allocation. The outperformance may be attributed to the the multi-period planning (versus single-period planning) and the data-driven RL approach (versus classical estimation approach).

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.
The authors propose an incremental step toward combining the insights of modern portfolio theory with some of the propensities documented in the literature on behavioral finance. They develop a goals-based wealth management approach that finds a specific subportfolio to address each of an investor's goals and then derive the least-cost solution. T hey relate the closed-form solution for the one-period, two-asset problem to the mean-variance efficient frontier. Consistent with the - lockbox separation- concept proposed by Sharpe, they demonstrate that a multiperiod goal, such as a retirement plan, can be viewed as a collection of single-period problems. Next, they extend their result to a market with many assets, where portfolios are exogenously given. Finally, they illustrate the approach with a case study with multiple asset classes and multiperiod goals.

Wang, T., Bao, X., Clavera, I., Hoang, J., Wen, Y., Langlois, E., Zhang, S., Zhang, G., Abbeel, P., and Ba, J. (2019). "Benchmarking Model-Based Reinforcement Learning." In: *arXiv e-Print*.
Model-based reinforcement learning (MBRL) is widely seen as having the potential to be significantly more sample efficient than model-free RL. However, research in model-based RL has not been very standardized. It is fairly common for authors to experiment with self-designed environments, and there are several separate lines of research, which are sometimes closed-sourced or not reproducible. Accordingly, it is an open question how these various existing MBRL algorithms perform relative to each other. To facilitate research in MBRL, in this paper we gather a wide collection of MBRL algorithms and propose over 18 benchmarking environments specially designed for MBRL. We benchmark these algorithms with unified problem settings, including noisy environments. Beyond cataloguing performance, we explore and unify the underlying algorithmic differences across MBRL algorithms. We characterize three key research challenges for future MBRL research: the dynamics bottleneck, the planning horizon dilemma, and the early-termination dilemma. Finally, to maximally facilitate future research on MBRL, we open-source our benchmark in `http://www.cs.toronto.edu/~tingwuwang/mbrl.html`.

Wells, L. and Bednarz, T. (2021). "Explainable AI and Reinforcement Learning—A Systematic Review of Current Approaches and Trends." In: *Frontiers in Artificial Intelligence* 4.
Research into Explainable Artificial Intelligence (XAI) has been increasing in recent years as a response to the need for increased transparency and trust in AI. This is particularly important as AI is used in sensitive domains with societal, ethical, and safety implications. Work in XAI has primarily focused on Machine Learning (ML) for classification, decision, or action, with detailed systematic reviews already undertaken. This review looks to explore current approaches and limitations for XAI in the area of Reinforcement Learning (RL). From 520 search results, 25 studies (including 5 snowball sampled) are reviewed, highlighting visualization, query-based explanations, policy summarization, human-in-the-loop collaboration, and verification as trends in this area. Limitations in the studies are presented, particularly a lack of user studies, and the prevalence of toy-examples and difficulties providing understandable explanations. Areas for future study are identified, including immersive visualization, and symbolic representation.

Weng, J., Chen, H., Yan, D., You, K., Duburcq, A., Zhang, M., Su, H., and Zhu, J. (2022). "Tianshou: a Highly Modularized Deep Reinforcement Learning Library." In: *arXiv e-Print*.
We present Tianshou, a highly modularized python library for deep reinforcement learning (DRL) that uses PyTorch as its backend. Tianshou aims to provide building blocks to replicate common RL experiments and has officially supported more than 15 classic algorithms succinctly. To facilitate related research and prove Tianshou's reliability, we release Tianshou's benchmark of MuJoCo environments, covering 9 classic algorithms and 9/13 Mujoco tasks with state-of-the-art performance. We open-sourced Tianshou at `https:`

//github.com/thu-ml/tianshou, which has received over 3k stars and become one of the most popular PyTorch-based DRL libraries.

Wiecki, T., Campbell, A., Lent, J., and Stauth, J. (2016). "All That Glitters Is Not Gold: Comparing Backtest and Out-of-Sample Performance on a Large Cohort of Trading Algorithms." In: *The Journal of Investing* 25(3), pp. 69–80.

When automated trading strategies are developed and evaluated using backtests on historical pricing data, there exists a tendency to overfit to the past. Using a unique dataset of 888 algorithmic trading strategies developed and backtested on the Quantopian platform, with at least six months of out-of-sample performance, this article studies the prevalence and impact of backtest overfitting. Specifically, the authors find that commonly reported backtest evaluation metrics, such as the Sharpe ratio, offer little value in predicting out-of-sample performance ($R2 < 0.025$). In contrast, higher-order moments, such as volatility and maximum drawdown, as well as portfolio construction features (e.g., hedging), show significant predictive value of relevance to quantitative finance practitioners. Moreover, in line with prior theoretical considerations, the authors find empirical evidence of overfitting-the more backtesting a quant has done for a strategy, the larger the discrepancy between backtest and out-of-sample performance. Finally, they show that by training nonlinear, machine-learning classifiers on a variety of features that describe backtest behavior, out-of-sample performance can be predicted with much greater accuracy ($R2 = 0.17$) on hold-out data than when using linear, univariate features. A portfolio constructed by using predictions on hold-out data performed significantly better out-of-sample than one constructed from algorithms with the highest backtest Sharpe ratios.

Xing, L. (2019). "Learning and Exploiting Multiple Subgoals for Fast Exploration in Hierarchical Reinforcement Learning." In: *arXiv e-Print*.

Hierarchical Reinforcement Learning (HRL) exploits temporally extended actions, or options, to make decisions from a higher-dimensional perspective to alleviate the sparse reward problem, one of the most challenging problems in reinforcement learning. The majority of existing HRL algorithms require either significant manual design with respect to the specific environment or enormous exploration to automatically learn options from data. To achieve fast exploration without using manual design, we devise a multi-goal HRL algorithm, consisting of a high-level policy Manager and a low-level policy Worker. The Manager provides the Worker multiple subgoals at each time step. Each subgoal corresponds to an option to control the environment. Although the agent may show some confusion at the beginning of training since it is guided by three diverse subgoals, the agent's behavior policy will quickly learn how to respond to multiple subgoals from the high-level controller on different occasions. By exploiting multiple subgoals, the exploration efficiency is significantly improved. We conduct experiments in Atari's Montezuma's Revenge environment, a well-known sparse reward environment, and in doing so achieve the same performance as state-of-the-art HRL methods with substantially reduced training time cost.

Xiong, Z., Liu, X.-Y., Zhong, S., Yang, H., and Walid, A. (2018). "Practical Deep Reinforcement Learning Approach for Stock Trading." In: *arXiv e-Print*.

Stock trading strategy plays a crucial role in investment companies. However, it is challenging to obtain optimal strategy in the complex and dynamic stock market. We explore the potential of deep reinforcement learning to optimize stock trading strategy and thus maximize investment return. 30 stocks are selected as our trading stocks and their daily prices are used as the training and trading market environment. We train a deep reinforcement learning agent and obtain an adaptive trading strategy. The agent's performance is evaluated and compared with Dow Jones Industrial Average and the traditional min-variance portfolio allocation strategy. The proposed deep reinforcement learning approach is shown to outperform the two baselines in terms of both the Sharpe ratio and cumulative returns.

Xu, R. and Chen, Z. (2021). "A Validation Tool for Designing Reinforcement Learning Environments." In: *arXiv e-Print*.

Reinforcement learning (RL) has gained increasing attraction in the academia and tech industry with launches to a variety of impactful applications and products. Although research is being actively conducted on many fronts (e.g., offline RL, performance, etc.), many RL practitioners face a challenge that has been largely ignored: determine whether a designed Markov Decision Process (MDP) is valid and meaningful. This study proposes a heuristic-based feature analysis method to validate whether an MDP is well formulated. We believe an MDP suitable for applying RL should contain a set of state features that are both sensitive to actions and predictive in rewards. We tested our method in constructed environments showing that our approach can identify certain invalid environment formulations. As far as we know, performing validity analysis for RL

problem formulation is a novel direction. We envision that our tool will serve as a motivational example to help practitioners apply RL in real-world problems more easily.

Yaghmaie, F. A. and Ljung, L. (2021). "A Crash Course on Reinforcement Learning." In: *arXiv e-Print*.
The emerging field of Reinforcement Learning (RL) has led to impressive results in varied domains like strategy games, robotics, etc. This handout aims to give a simple introduction to RL from control perspective and discuss three possible approaches to solve an RL problem: Policy Gradient, Policy Iteration, and Model-building. Dynamical systems might have discrete action-space like cartpole where two possible actions are +1 and -1 or continuous action space like linear Gaussian systems. Our discussion covers both cases.

Yang, H., Liu, X.-Y., Zhong, S., and Walid, A. (2020). "Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy." In: *SSRN e-Print*.
Stock trading strategies play a critical role in investment. However, it is challenging to design a profitable strategy in a complex and dynamic stock market. In this paper, we propose an ensemble strategy that employs deep reinforcement schemes to learn a stock trading strategy by maximizing investment return. We train a deep reinforcement learning agent and obtain an ensemble trading strategy using three actor-critic based algorithms: Proximal Policy Optimization (PPO), Advantage Actor Critic (A2C), and Deep Deterministic Policy Gradient (DDPG). The ensemble strategy inherits and integrates the best features of the three algorithms, thereby robustly adjusting to different market situations. In order to avoid the large memory consumption in training networks with continuous action space, we employ a load-on-demand technique for processing very large data. We test our algorithms on the 30 Dow Jones stocks that have adequate liquidity. The performance of the trading agent with different reinforcement learning algorithms is evaluated and compared with both the Dow Jones Industrial Average index and the traditional min-variance portfolio allocation strategy. The proposed deep ensemble strategy is shown to outperform the three individual algorithms and two baselines in terms of the risk-adjusted return measured by the Sharpe ratio.

Yang, T., Tang, H., Bai, C., Liu, J., Hao, J., Meng, Z., and Liu, P. (2022). "Exploration in Deep Reinforcement Learning: A Comprehensive Survey." In: *arXiv e-Print*.
Deep Reinforcement Learning (DRL) and Deep Multi-agent Reinforcement Learning (MARL) have achieved significant success across a wide range of domains, such as game AI, autonomous vehicles, robotics and finance. However, DRL and deep MARL agents are widely known to be sample-inefficient and millions of interactions are usually needed even for relatively simple game settings, thus preventing the wide application in real-industry scenarios. One bottleneck challenge behind is the well-known exploration problem, i.e., how to efficiently explore the unknown environments and collect informative experiences that could benefit the policy learning most. In this paper, we conduct a comprehensive survey on existing exploration methods in DRL and deep MARL for the purpose of providing understandings and insights on the critical problems and solutions. We first identify several key challenges to achieve efficient exploration, which most of the exploration methods aim at addressing. Then we provide a systematic survey of existing approaches by classifying them into two major categories: uncertainty-oriented exploration and intrinsic motivation-oriented exploration. The essence of uncertainty-oriented exploration is to leverage the quantification of the epistemic and aleatoric uncertainty to derive efficient exploration. By contrast, intrinsic motivation-oriented exploration methods usually incorporate different reward agnostic information for intrinsic exploration guidance. Beyond the above two main branches, we also conclude other exploration methods which adopt sophisticated techniques but are difficult to be classified into the above two categories. In addition, we provide a comprehensive empirical comparison of exploration methods for DRL on a set of commonly used benchmarks. Finally, we summarize the open problems of exploration in DRL and deep MARL and point out a few future directions.

Yashaswi, K. (2021). "Deep Reinforcement Learning for Portfolio Optimization using Latent Feature State Space (LFSS) Module." In: *arXiv e-Print*.
Dynamic Portfolio optimization is the process of distribution and rebalancing of a fund into different financial assets such as stocks, cryptocurrencies, etc, in consecutive trading periods to maximize accumulated profits or minimize risks over a time horizon. This field saw huge developments in recent years, because of the increased computational power and increased research in sequential decision making through control theory. Recently Reinforcement Learning(RL) has been an important tool in the development of sequential and dynamic portfolio optimization theory. In this paper, we design a Deep Reinforcement Learning(DRL) framework as an autonomous portfolio optimization agent consisting of a Latent Feature State Space(LFSS) Module for filtering and feature extraction of financial data which is used as a state space for deep RL model. We develop an extensive RL agent with high efficiency and performance advantages over several benchmarks and

model-free RL agents used in prior work. The noisy and non-stationary behaviour of daily asset prices in the financial market is addressed through Kalman Filter. Autoencoders, ZoomSVD, and restricted Boltzmann machines were the models used and compared in the module to extract relevant time series features as state space. We simulate weekly data, with practical constraints and transaction costs, on a portfolio of S&P 500 stocks. We introduce a new benchmark based on technical indicator Kd-Index and Mean-Variance Model as compared to equal weighted portfolio used in most of the prior work. The study confirms that the proposed RL portfolio agent with state space function in the form of LFSS module gives robust results with an attractive performance profile over baseline RL agents and given benchmarks.

Yekelchik, B., Tatsat, H., and Coriarty, Z. (2021). "Deep Q-Network Interpertability: Applications to ETF Trading." In: *SSRN e-Print*.

We present an interpertability infrastructure for reinforcement learning (RL) based trading strategies. For all audiences to be able to answer the the question of 'how does the algorithm work?', we provide a visual and user-friendly approach, in contrast to a more quantitative approach. This allows not only a technical audience to consume insights derived from an RL-based trading approach. In this application, we introduce a three module approach in understanding value-based RL, specifically Deep Q-Learning. We demonstrate this infrastructure and possible derived outcomes of using this infrastructrure when applied to trading a market ETF in a given time interval.

Yu, L. (2021). "Comparing Classical Portfolio Optimization and Robust Portfolio Optimization on Black Swan Events." MA thesis. University of Waterloo.

Black swan events, such as natural catastrophes and manmade market crashes, historically have a drastic negative influence on investments; and there is a discrepancy on losses caused by these two types of disasters. In general, there is a recovery and it is of interest to understand what type of investment strategies lead to better performance for investors. In this thesis we study classical portfolio optimization, robust portfolio optimization and some historical black swan events. We compare two main strategies: mean variance optimization vs robust portfolio optimization on two types of black swan events: natural vs anthropogenic. The comparison illustrates that robust portfolio optimization is much more conservative, and has a shorter recovery time than classical portfolio optimization. Moreover, the losses in the stock investment resulted from a natural disaster are very minor compared to the losses resulted from an anthropogenic market crash.

Yuan, M. and Zhou, G. (2022). "Why Naive 1/N Diversification Is Not So Naive, and How to Beat It?" In: *SSRN e-Print*.

In this paper, we study portfolio choice problem under estimation risk and show why the 1/N rule is very difficult to beat in applications and studies. First, as long as the dimensionality is high relative to sample size, we show that the usual estimated investment strategies are biased even asymptotically. Second, we show that the 1/N rule is optimal in a one-factor model with diversifiable risks as dimensionality increases, irrespectively of the sample size, making investment theory-based rules inadequate as they suffer from estimation errors. Third, we provide strategies that can outperform the 1/N under suitable conditions.

Zhang, B., Rajan, R., Pineda, L., Lambert, N., Biedenkapp, A., Chua, K., Hutter, F., and Calandra, R. (2021). "On the Importance of Hyperparameter Optimization for Model-based Reinforcement Learning." In: *arXiv e-Print*.

Model-based Reinforcement Learning (MBRL) is a promising framework for learning control in a data-efficient manner. MBRL algorithms can be fairly complex due to the separate dynamics modeling and the subsequent planning algorithm, and as a result, they often possess tens of hyperparameters and architectural choices. For this reason, MBRL typically requires significant human expertise before it can be applied to new problems and domains. To alleviate this problem, we propose to use automatic hyperparameter optimization (HPO). We demonstrate that this problem can be tackled effectively with automated HPO, which we demonstrate to yield significantly improved performance compared to human experts. In addition, we show that tuning of several MBRL hyperparameters dynamically, i.e. during the training itself, further improves the performance compared to using static hyperparameters which are kept fixed for the whole training. Finally, our experiments provide valuable insights into the effects of several hyperparameters, such as plan horizon or learning rate and their influence on the stability of training and resulting rewards.

Zhang, C., Vinyals, O., Munos, R., and Bengio, S. (2018). "A Study on Overfitting in Deep Reinforcement Learning." In: *arXiv e-Print*.

Recent years have witnessed significant progresses in deep Reinforcement Learning (RL). Empowered with large scale neural networks, carefully designed architectures, novel training algorithms and massively parallel

computing devices, researchers are able to attack many challenging RL problems. However, in machine learning, more training power comes with a potential risk of more overfitting. As deep RL techniques are being applied to critical problems such as healthcare and finance, it is important to understand the generalization behaviors of the trained agents. In this paper, we conduct a systematic study of standard RL agents and find that they could overfit in various ways. Moreover, overfitting could happen "robustly": commonly used techniques in RL that add stochasticity do not necessarily prevent or detect overfitting. In particular, the same agents and learning algorithms could have drastically different test performance, even when all of them achieve optimal rewards during training. The observations call for more principled and careful evaluation protocols in RL. We conclude with a general discussion on overfitting in RL and a study of the generalization behaviors from the perspective of inductive bias.

Zhang, C., Li, Y., Chen, X., Jin, Y., Tang, P., and Li, J. (2020a). "DoubleEnsemble: A New Ensemble Method Based on Sample Reweighting and Feature Selection for Financial Data Analysis." In: *IEEE International Conference on Data Mining (ICDM)*. IEEE.

Modern machine learning models (such as deep neural networks and boosting decision tree models) have become increasingly popular in financial market prediction, due to their superior capacity to extract complex non-linear patterns. However, since financial datasets have very low signal-to-noise ratio and are non-stationary, complex models are often very prone to overfitting and suffer from instability issues. Moreover, as various machine learning and data mining tools become more widely used in quantitative trading, many trading firms have been producing an increasing number of features (aka factors). Therefore, how to automatically select effective features becomes an imminent problem. To address these issues, we propose DoubleEnsemble, an ensemble framework leveraging learning trajectory based sample reweighting and shuffling based feature selection. Specifically, we identify the key samples based on the training dynamics on each sample and elicit key features based on the ablation impact of each feature via shuffling. Our model is applicable to a wide range of base models, capable of extracting complex patterns, while mitigating the overfitting and instability issues for financial market prediction. We conduct extensive experiments, including price prediction for cryptocurrencies and stock trading, using both DNN and gradient boosting decision tree as base models. Our experiment results demonstrate that DoubleEnsemble achieves a superior performance compared with several baseline methods.

Zhang, F., Guo, R., and Cao, H. (2020b). "Information Coefficient as a Performance Measure of Stock Selection Models." In: *arXiv e-Print*.

Information coefficient (IC) is a widely used metric for measuring investment managers' skills in selecting stocks. However, its adequacy and effectiveness for evaluating stock selection models has not been clearly understood, as IC from a realistic stock selection model can hardly be materially different from zero and is often accompanies with high volatility. In this paper, we investigate the behavior of IC as a performance measure of stick selection models. Through simulation and simple statistical modeling, we examine the IC behavior both statically and dynamically. The examination helps us propose two practical procedures that one may use for IC-based ongoing performance monitoring of stock selection models.

Zhang, H. and Yu, T. (2020). "Taxonomy of Reinforcement Learning Algorithms." In: *Deep Reinforcement Learning*. Springer Singapore, pp. 125–133.

In this chapter, we introduce and summarize the taxonomy and categories for reinforcement learning (RL) algorithms. Figure 3.1 presents an overview of the typical and popular algorithms in a structural way. We classify reinforcement learning algorithms from different perspectives, including model-based and model-free methods, value-based and policy-based methods (or combination of the two), Monte Carlo methods and temporal-difference methods, on-policy and off-policy methods. Most reinforcement learning algorithms can be classified under different categories according to the above criteria, hope this helps to provide the readers some overviews of the full picture before introducing the algorithms in detail in later chapters.

Zhang, Z., Zohren, S., and Roberts, S. (2020c). "Deep Learning for Portfolio Optimization." In: *The Journal of Financial Data Science* 22(4), pp. 8–20.

In this article, the authors adopt deep learning models to directly optimize the portfolio Sharpe ratio. The framework they present circumvents the requirements for forecasting expected returns and allows them to directly optimize portfolio weights by updating model parameters. Instead of selecting individual assets, they trade exchange-traded funds of market indexes to form a portfolio. Indexes of different asset classes show robust correlations, and trading them substantially reduces the spectrum of available assets from which to choose. The authors compare their method with a wide range of algorithms, with results showing that the

model obtains the best performance over the testing period of 2011 to the end of April 2020, including the financial instabilities of the first quarter of 2020. A sensitivity analysis is included to clarify the relevance of input features, and the authors further study the performance of their approach under different cost rates and different risk levels via volatility scaling.

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*.
The recommender system is an important form of intelligent application, which assists users to alleviate from information redundancy. Among the metrics used to evaluate a recommender system, the metric of conversion has become more and more important. The majority of existing recommender systems perform poorly on the metric of conversion due to its extremely sparse feedback signal. To tackle this challenge, we propose a deep hierarchical reinforcement learning based recommendation framework, which consists of two components, i.e., high-level agent and low-level agent. The high-level agent catches long-term sparse conversion signals, and automatically sets abstract goals for low-level agent, while the low-level agent follows the abstract goals and interacts with real-time environment. To solve the inherent problem in hierarchical reinforcement learning, we propose a novel deep hierarchical reinforcement learning algorithm via multi-goals abstraction (HRL-MG). Our proposed algorithm contains three characteristics: 1) the high-level agent generates multiple goals to guide the low-level agent in different stages, which reduces the difficulty of approaching high-level goals; 2) different goals share the same state encoder parameters, which increases the update frequency of the high-level agent and thus accelerates the convergence of our proposed algorithm; 3) an appreciate benefit assignment function is designed to allocate rewards in each goal so as to coordinate different goals in a consistent direction. We evaluate our proposed algorithm based on a real-world e-commerce dataset and validate its effectiveness.

Zhu, Z., Lin, K., and Zhou, J. (2021). "Transfer Learning in Deep Reinforcement Learning: A Survey." In: *arXiv e-Print*.
This paper surveys the field of transfer learning in the problem setting of Reinforcement Learning (RL). RL has been a key solution to sequential decision-making problems. Along with the fast advances of RL in various domains, such as robotics and game-playing, transfer learning arises as an important technique to assist RL by leveraging and transferring external expertise to boost the learning process of RL. In this survey, we review the central issues of transfer learning in the RL domain, providing a systematic categorization of its state-of-the-art techniques. We analyze their goals, methodologies, applications, and the RL frameworks under which the transfer learning techniques are approachable. We discuss the relationship between transfer learning and other relevant topics from the RL perspective and also explore the potential challenges as well as future development directions for transfer learning in RL.