Applying Partial Observable Markov Decision Processes to Trading: A Practical Approach

Omniscience Research
Unregistered User

Abstract

The application of Partial Observable Markov Decision Processes (POMDPs) to trading strategies presents a compelling approach to managing uncertainty in financial markets. This paper explores the practical implementation of POMDPs within the trading domain, focusing on real-world applications rather than theoretical models. We begin by reviewing the foundational concepts of POMDPs, including their definition, components, and solution methods. We then discuss the adaptation of POMDPs to the trading environment, highlighting the unique challenges posed by financial markets. Through case studies, we demonstrate the successful application of POMDPs to trading strategies and compare their performance to traditional models. Our analysis extends to the scalability and robustness of POMDP-based systems, addressing the handling of high-dimensional spaces and market anomalies. Finally, we consider the future potential of POMDPs in trading, in light of advancements in computational power and machine learning. The paper aims to provide a comprehensive overview of POMDPs in trading, offering insights for both practitioners and researchers in the field of finance.

1 Background on POMDPs

Partial Observable Markov Decision Processes (POMDPs) provide a mathematical framework for decision-making in environments with uncertainty and incomplete information. They extend the Markov Decision Process (MDP) model by incorporating the element of partial observability, which is particularly relevant in the context of financial trading where market states are not fully observable.

1.1 Definition and Components of a POMDP

A POMDP is defined by a tuple $(S, A, T, R, \Omega, O, \gamma)$, where S is a finite set of states, A is a finite set of actions, $T : S \times A \times S \to [0, 1]$ is the state transition probability function, $R : S \times A \to \mathbb{R}$ is the reward function, Ω is a finite set of observations, $O : S \times A \times \Omega \to [0, 1]$ is the observation probability function, and $\gamma \in [0, 1)$ is the discount factor Kaelbling et al. [1998].

The key challenge in POMDPs is that the agent does not have direct access to the underlying state. Instead, it must make decisions based on a belief state, which is a probability distribution over the possible states, updated as new observations are made Shani et al. [2013].

1.2 Solving POMDPs: Methods and Algorithms

Solving a POMDP involves finding a policy that maximizes the expected cumulative discounted reward. This is challenging due to the need to maintain and update a belief state. Exact solutions can be computed using value iteration algorithms, which iteratively update the value of each belief state until convergence Monahan [1982]. However, these methods are computationally intractable for large state spaces.

Preprint. Under review.

Approximate methods, such as point-based value iteration (PBVI), overcome this by focusing on a finite set of representative belief points, significantly reducing computational requirements Pineau et al. [2003]. Other approaches include policy gradient methods, which optimize a parameterized policy directly Sutton et al. [2000], and model-based reinforcement learning, which attempts to learn the POMDP model components from interaction with the environment Ross et al. [2008].

1.3 Advantages of POMDPs in Decision-Making Under Uncertainty

POMDPs are well-suited for environments where the agent must make a sequence of decisions without perfect information. They provide a principled approach to handling uncertainty, allowing the agent to reason about the possible states of the world and the likelihood of different outcomes. This is particularly advantageous in trading, where market conditions can change rapidly and unpredictably Bäuerle and Rieder [2011].

Moreover, POMDPs can incorporate risk preferences by adjusting the reward function or the belief update process, enabling the creation of trading strategies that align with the risk tolerance of the investor Howard and Matheson [1972]. This flexibility makes POMDPs a powerful tool for developing sophisticated trading algorithms that can adapt to the complexities of financial markets.

In summary, POMDPs offer a robust framework for decision-making in uncertain environments, with a range of solution methods that can be tailored to the specific requirements of trading applications. Their ability to handle incomplete information and adapt to changing conditions makes them an attractive choice for developing advanced trading strategies.

2 POMDPs in the Context of Trading

The application of Partial Observable Markov Decision Processes (POMDPs) to trading is a natural extension of their use in other domains where decision-making under uncertainty is crucial. In trading, the uncertainty arises from the stochastic nature of financial markets, where the true state of the market is never fully known and is influenced by a myriad of factors, including economic indicators, political events, and trader psychology.

2.1 Mapping Trading Problems to POMDP Frameworks

To apply POMDPs to trading, one must first establish a correspondence between the elements of a POMDP and the components of a trading system. The state space S can represent various market conditions or regimes, such as trending, mean-reverting, or high-volatility periods. The action space A typically includes buying, selling, or holding an asset, as well as more complex actions like adjusting stop-loss or take-profit levels. The transition probabilities T capture the likelihood of moving from one market state to another, given an action. These probabilities are often estimated using historical data or through simulation techniques [Brockwell and Davis, 2002].

Observations in a trading POMDP can be derived from market data such as price movements, volume, or derived technical indicators. The observation function O relates the underlying market state to these observable quantities, accounting for noise and other sources of uncertainty. The reward function R is designed to reflect the trader's objectives, such as profit maximization or drawdown minimization, and can be adjusted to incorporate transaction costs and risk preferences [Moody and Saffell, 2001].

2.2 Unique Aspects of Trading that Affect POMDP Modeling

Trading environments present several unique challenges for POMDP modeling. First, financial markets are non-stationary; the statistical properties of market data can change over time, which can invalidate the assumptions underlying the transition and observation models [Tsay, 2005]. This non-stationarity requires adaptive algorithms that can update the POMDP model as new data becomes available.

Second, the action space in trading is often constrained by regulations, liquidity, and capital limitations. These constraints must be incorporated into the POMDP model to ensure that the resulting policies are feasible in practice [Almgren and Chriss, 2001].

Third, the reward structure in trading is asymmetric due to the presence of transaction costs and the fact that losses can have a disproportionate impact on the trader's capital compared to gains. This asymmetry must be carefully considered when designing the reward function to ensure that the POMDP policy does not take excessive risks [Cont, 2001].

Lastly, the financial markets are influenced by the actions of numerous participants, leading to complex dynamics that can be difficult to capture in a POMDP model. Strategic behavior, such as the impact of large trades on market prices, must be considered to avoid adverse effects on the performance of the trading strategy [Cartea et al., 2015].

In addressing these challenges, POMDPs offer a flexible and powerful framework for developing trading strategies that can navigate the complexities of financial markets. By explicitly modeling uncertainty and adapting to changing market conditions, POMDP-based trading systems have the potential to outperform traditional approaches that do not account for these factors.

The successful application of POMDPs to trading hinges on the careful consideration of market characteristics and the trader's objectives. By capturing the nuances of the trading environment within the POMDP framework, researchers and practitioners can develop strategies that are both robust and responsive to the ever-evolving landscape of financial markets.

3 Data Representation and Preprocessing

The success of POMDP-based trading strategies is heavily dependent on the quality and representation of the underlying data. Financial markets generate vast amounts of data, which must be processed and transformed into a format suitable for POMDP modeling. This section discusses the sources of trading data and the preprocessing techniques necessary to facilitate the application of POMDPs in trading.

3.1 Sources of Trading Data

Trading data can be broadly categorized into two types: historical and real-time. Historical data encompasses past records of market prices, volumes, and other financial indicators, which are essential for backtesting and training POMDP models Brownlees and Gallo [2010]. Real-time data, on the other hand, is used for live trading and includes streaming price quotes, order book dynamics, and transaction feeds Hasbrouck [2007].

In addition to market data, alternative data sources such as news articles, social media sentiment, and economic reports can provide valuable insights into market conditions and trader behavior Tetlock [2007]. These alternative data sources often require natural language processing and sentiment analysis techniques to extract relevant features for POMDP models [Bollen et al., 2011].

3.2 Techniques for Transforming Raw Data into POMDP-Compatible Formats

Preprocessing trading data for POMDPs involves several steps, each designed to enhance the model's ability to capture the underlying market dynamics and make informed decisions.

3.2.1 Data Cleaning

The initial step in data preprocessing is cleaning, which involves removing errors, duplicates, and outliers from the dataset. This step is crucial as erroneous data can lead to misleading conclusions and suboptimal trading decisions Brown and Jennings [1989]. Techniques such as filtering based on volume thresholds and statistical outlier detection are commonly employed to ensure data integrity Zhang and Zhang [2005].

3.2.2 Feature Engineering

Feature engineering is the process of transforming raw data into a set of features that POMDP models can utilize. In trading, features may include technical indicators like moving averages, oscillators, and momentum measures, which provide a condensed view of market trends and volatility Murphy [1999]. Additionally, dimensionality reduction techniques such as principal component analysis (PCA) can be used to identify the most informative features and reduce the computational complexity of the POMDP model Jolliffe [2016].

3.2.3 Normalization and Scaling

Financial data often exhibit significant variations in scale and distribution, which can affect the performance of learning algorithms. Normalization and scaling techniques, such as min-max scaling or z-score standardization, are applied to ensure that all features contribute equally to the POMDP's decision-making process Patro and Sahu [2015].

3.2.4 Temporal Alignment

Temporal alignment is necessary when combining data from different sources or frequencies. This involves resampling and interpolating data to match a common time frame, ensuring that the POMDP model receives a coherent and synchronized view of the market Brownlees and Gallo [2010].

3.2.5 Sequence Generation

POMDPs require sequences of observations to infer the hidden state of the market. Sliding window techniques are used to create overlapping sequences of data points, which serve as input to the POMDP model. The length of the window is a critical parameter that balances the trade-off between the recency of information and the context provided by longer historical sequences De Prado [2012].

By meticulously preprocessing trading data, researchers and practitioners can construct POMDP models that are well-equipped to navigate the complexities of financial markets. The transformation of raw data into a structured and informative format is a foundational step in the development of POMDP-based trading strategies, setting the stage for robust and intelligent decision-making in the face of uncertainty.

4 Model Specification for Trading

The specification of a POMDP model for trading involves defining the states, actions, observations, and reward structure. These components must be carefully designed to reflect the complexities of financial markets and the objectives of the trading strategy.

4.1 Defining States in Trading POMDPs

The state in a POMDP represents the underlying condition of the system, which in the context of trading, corresponds to the market regime or the financial instrument's price dynamics. However, the true state of the market is not directly observable, and traders must infer it from indirect observations. Common approaches to state representation include discretizing price ranges, using latent variables from statistical models such as Hidden Markov Models (HMMs), or employing market indicators that signal different market conditions [Hassan and Nath, 2005,R].

4.1.1 Market Regime Identification

Market regimes can be characterized by various phases such as trending, mean-reverting, or high-volatility periods. Identifying these regimes is crucial for adapting trading strategies to current market conditions. Techniques such as regime-switching models can be used to infer the hidden states that correspond to different market regimes [Hamilton, 1989].

4.1.2 Financial Instrument Dynamics

For individual financial instruments, states may represent specific price movement patterns or levels of liquidity. These states can be modeled using time-series analysis techniques that capture the temporal dependencies in price changes and trading volumes [Tsay, 2005].

4.2 Actions and Action Space

In a trading POMDP, actions correspond to the possible trades or orders that can be executed, such as buy, sell, or hold positions. The action space can be discrete, with a finite set of actions, or continuous, allowing for a range of trade sizes or prices. The design of the action space must consider the trade-off between granularity and computational tractability [Nevmyvaka et al., 2006].

4.2.1 Discrete Action Space

A discrete action space simplifies the decision-making process by limiting the number of possible actions. This approach is suitable for strategies that require quick decisions and can tolerate a coarse level of control over trade execution [Brockwell and Davis, 2002].

4.2.2 Continuous Action Space

A continuous action space provides a finer level of control, enabling precise trade sizes and prices. This is particularly important for strategies that aim to optimize execution costs or manage large positions without significantly impacting the market [Almgren and Chriss, 2001].

4.3 Observations and Observation Space

Observations in a trading POMDP are the pieces of information available to the trader, which are used to update beliefs about the current state. These may include price quotes, trade volumes, bid-ask spreads, and other market data. The observation space must be designed to capture the relevant information while avoiding information overload [Madhavan, 2002].

4.3.1 Market Data

Market data observations are the most direct source of information, reflecting the latest market activity. These observations are typically noisy and may contain artifacts of market microstructure, which must be accounted for in the POMDP model [O'Hara, 1995].

4.3.2 Derived Observations

Derived observations include features extracted from raw market data, such as technical indicators or outputs from statistical models. These observations can provide a more structured and informative view of the market, aiding in the state inference process Murphy [1999].

4.4 Reward Structures Tailored to Trading Objectives

The reward structure in a trading POMDP defines the objectives of the trading strategy. It translates desired outcomes, such as profit maximization or risk minimization, into numerical rewards that guide the learning and decision-making process [Moody and Saffell, 2001].

4.4.1 Profit-Based Rewards

Profit-based rewards are straightforward and align directly with the goal of financial gain. They can be defined as the change in portfolio value or the return on investment from executed trades [Sharpe, 1994].

4.4.2 Risk-Adjusted Rewards

Risk-adjusted rewards incorporate measures of risk to balance the pursuit of profits with the management of potential losses. Metrics such as the Sharpe ratio or Value at Risk (VaR) can be integrated into the reward function to encourage strategies that achieve favorable risk-return profiles [Sharpe, 1994,J].

By meticulously crafting the components of the POMDP model, traders can create sophisticated strategies that are capable of navigating the intricate dance of the markets. The careful consideration of states, actions, observations, and rewards ensures that the POMDP framework is not only theoretically sound but also practically effective in the relentless pursuit of alpha.

5 Learning and Adaptation in POMDPs

The dynamic nature of financial markets necessitates that trading strategies adapt over time to maintain their effectiveness. Learning and adaptation mechanisms within POMDPs are critical for updating the model as new information becomes available and market conditions evolve.

5.1 Online Learning Algorithms for Trading POMDPs

Online learning in the context of POMDPs refers to the continuous updating of the model's parameters and policy based on incoming observations and rewards. This is essential in trading, where the model must adapt to real-time market information to make timely decisions.

5.1.1 Reinforcement Learning Approaches

Reinforcement learning (RL) algorithms, such as Q-learning and policy gradient methods, have been successfully applied to POMDPs in trading. These algorithms iteratively improve the trading policy by learning from the outcomes of past actions [Sutton and Barto, 2018]. For instance, Q-learning can be used to estimate the optimal action-value function, which represents the expected utility of taking a given action in a given state and following the optimal policy thereafter [Watkins and Dayan, 1992].

5.1.2 Bayesian Methods

Bayesian methods provide a probabilistic approach to learning in POMDPs, where beliefs about the state are updated using Bayes' theorem. In trading, Bayesian methods can incorporate the uncertainty in market predictions and adapt to changes in the underlying data-generating process [Ghavamzadeh et al., 2015]. The use of particle filters or sequential Monte Carlo methods allows for the approximation of the belief state, which is particularly useful when dealing with high-dimensional or continuous state spaces [Doucet et al., 2000].

5.2 Adaptive Mechanisms to Cope with Market Changes

Adaptation in trading POMDPs involves mechanisms that allow the model to adjust to structural breaks or regime shifts in the market. These mechanisms must be responsive to changes while avoiding overfitting to noise or transient market conditions.

5.2.1 Regime-Switching Models

Regime-switching models enable the POMDP to recognize and adapt to different market regimes, such as shifts from bull to bear markets or periods of high volatility. By incorporating multiple models, each tailored to a specific regime, the POMDP can switch between these models based on the inferred state of the market [Hamilton, 1990]. This approach allows for more flexible and robust trading strategies that can navigate diverse market conditions.

5.2.2 Meta-Learning

Meta-learning, or learning to learn, is an advanced adaptation technique where the learning algorithm itself is subject to optimization. In the context of trading POMDPs, meta-learning can be used to adjust the learning rate or to select the most appropriate model or algorithm based on performance feedback [Finn et al., 2017]. This level of adaptability is crucial for long-term success in trading, where the only constant is change.

The integration of learning and adaptation mechanisms into POMDP-based trading strategies is a testament to the ingenuity required to thrive in the financial markets. By harnessing the power of online learning and adaptive techniques, these strategies evolve in lockstep with the markets they seek to master, embodying the perpetual dance of analysis, prediction, and execution that defines the art of trading.

6 Case Studies of POMDPs in Trading

The theoretical underpinnings of POMDPs provide a robust framework for decision-making under uncertainty, but their practical effectiveness is best demonstrated through real-world applications. This section presents case studies where POMDPs have been successfully applied to trading strategies, highlighting the nuances of their implementation and the outcomes achieved.

6.1 Algorithmic Trading with POMDPs

Algorithmic trading strategies often involve complex decision-making in the face of uncertain market conditions. A study by Kearns and Nevmyvaka [Kearns et al., 2013] applied POMDPs to the problem of optimal execution, where the goal is to execute a large order with minimal market impact. They modeled the market as a POMDP, where the hidden state represented the true balance of supply and demand, and the observations were the recent history of prices and trades. Their approach used a reinforcement learning algorithm to learn the optimal policy for order execution, resulting in significant cost savings compared to standard execution strategies.

6.2 Portfolio Management with POMDPs

Portfolio management involves allocating assets to maximize returns while managing risk. A POMDP framework for portfolio management was proposed by Huang et al. [Huang et al., 2010], where the hidden state included economic indicators and the observations consisted of asset prices and market indices. The action space comprised different portfolio allocations, and the reward function was based on the portfolio's performance. By employing a point-based value iteration algorithm, the POMDP was able to outperform traditional mean-variance portfolio strategies, particularly in markets characterized by high uncertainty and rapid changes.

6.2.1 Adaptive Asset Allocation

In a more focused study, Shen and Si [Shen et al., 2014] explored adaptive asset allocation using a POMDP model that adjusted the portfolio in response to changing market conditions. They incorporated a regime-switching component to account for different market states, such as growth or recession. The POMDP dynamically rebalanced the portfolio by selecting from a predefined set of asset allocation strategies, each optimized for a specific market regime. This adaptive approach demonstrated improved risk-adjusted returns over static allocation strategies, showcasing the POMDP's ability to navigate complex, multi-regime environments.

The case studies presented in this section underscore the versatility and power of POMDPs in addressing the challenges of trading. By capturing the inherent uncertainty and providing a structured approach to learning and decision-making, POMDPs have carved out a niche in the sophisticated landscape of financial strategies. These real-world applications not only validate the theoretical appeal of POMDPs but also pave the way for their broader adoption in the trading community. As the financial markets continue to evolve, the fusion of POMDPs with trading acumen promises to yield strategies that are as dynamic and multifaceted as the markets themselves.

7 Performance Analysis

Evaluating the performance of POMDP-based trading systems is crucial to understanding their efficacy in real-world financial markets. This section dives into the criteria used for performance evaluation, the benchmarks set against traditional trading models, and the statistical significance of the results obtained from POMDP implementations.

7.1 Evaluation Criteria

The performance of trading systems is typically assessed using a variety of metrics that capture different aspects of trading effectiveness. For POMDP-based systems, the following criteria are commonly used:

- **Return on Investment (ROI)**: The ROI measures the net profit or loss made on the investments relative to the amount of money invested, providing a straightforward indicator of financial performance [Sharpe, 1994].
- **Sharpe Ratio**: This ratio adjusts the returns of an investment by the risk taken to achieve those returns, offering a risk-adjusted measure of performance Sortino and Price [1994].
- **Maximum Drawdown**: The maximum drawdown assesses the largest peak-to-trough decline in the value of a portfolio, reflecting the potential risk from a high loss Magdon-Ismail et al. [2004].
- **Calmar Ratio**: Similar to the Sharpe Ratio, the Calmar Ratio relates the annualized compound return to the maximum drawdown, emphasizing the trade-off between return and risk Young [1991].

These metrics provide a comprehensive view of the trading system's performance, taking into account both the returns generated and the risks incurred.

7.2 Benchmarking Against Traditional Models

To ascertain the value added by POMDP-based trading systems, it is essential to compare their performance with that of traditional trading models. Common benchmarks include:

- **Buy-and-Hold Strategy**: This strategy involves purchasing securities and holding them for a long period, serving as a baseline for comparison Fama and French [1988].
- **Mean-Variance Optimization**: A classical approach to portfolio allocation, mean-variance optimization aims to maximize returns for a given level of risk Markowitz [1952].
- **Technical Analysis Strategies**: These strategies use historical price and volume data to predict future market behavior Murphy [1999].

By comparing POMDP-based systems against these benchmarks, researchers can evaluate whether the complexity of POMDPs translates into superior performance.

7.2.1 Statistical Significance of Results

When analyzing the performance of POMDP-based trading systems, it is imperative to determine the statistical significance of the results. This involves using tests such as the t-test or the Mann-Whitney U test to compare the means of different samples Hollander et al. [2013]. Additionally, the bootstrap method can be employed to assess the stability of the performance metrics Efron and Tibshirani [1994]. These statistical methods ensure that the observed performance differences are not due to random chance but are indicative of the underlying efficacy of the POMDP approach.

The rigorous evaluation of POMDP-based trading systems through these criteria and benchmarks is a testament to the scientific approach adopted by researchers in this domain. By grounding their assessments in empirical evidence and statistical rigor, they provide a solid foundation for the continued exploration and refinement of POMDPs in the complex and ever-changing world of trading. The insights gleaned from such analyses not only inform the design of future systems but also contribute to the broader discourse on the role of advanced computational methods in financial decision-making.

8 Scalability and Robustness

The practical application of POMDPs in trading necessitates a thorough examination of their scalability and robustness. Financial markets are characterized by high-dimensional state spaces and rapid fluctuations, which pose significant challenges for POMDP models. This section explores the methods used to enhance the scalability of POMDPs to accommodate the vast array of financial instruments and the robustness required to withstand market anomalies and shocks.

8.1 Handling High-Dimensional Spaces

The dimensionality of state spaces in trading POMDPs can be immense due to the multitude of factors influencing market dynamics. Traditional POMDP solvers struggle with high-dimensional spaces due to the "curse of dimensionality" Bellman [1957]. To address this, several approaches have been proposed:

- **State Aggregation**: This technique involves grouping similar states together to reduce the effective size of the state space Singh and Sutton [1995].
- **Function Approximation**: Machine learning methods, such as neural networks, are used to approximate the value function over the state space, allowing for generalization across states [Sutton and Barto, 2018].
- **Monte Carlo Tree Search (MCTS)**: MCTS algorithms selectively sample the state space to efficiently estimate the value of actions without exhaustive enumeration Browne et al. [2012].

By leveraging these techniques, POMDP-based trading systems can scale to handle the complexity of financial markets, enabling them to make informed decisions across a broad spectrum of scenarios.

8.2 Ensuring Robustness

Robustness in trading systems is the ability to maintain performance despite unexpected market conditions or data anomalies. POMDPs must be designed to be resilient to such events to be viable for trading. The following strategies contribute to the robustness of POMDP-based trading systems:

8.2.1 Incorporating Risk Aversion

Risk-averse POMDP models incorporate preferences for lower variance in returns, which can be achieved by modifying the reward function to penalize high-risk actions Howard and Matheson [1972]. This modification often involves the use of utility functions that reflect the diminishing marginal utility of wealth, thereby naturally limiting exposure to high-risk positions.

8.2.2 Anomaly Detection and Response

Anomalies such as flash crashes or irregular market behavior can lead to significant losses if not properly managed. POMDPs can be equipped with anomaly detection mechanisms that trigger protective actions when irregular patterns are observed Kim [2014]. These mechanisms often rely on statistical models that identify deviations from normal market behavior, prompting the system to either exit positions or hedge against potential losses.

8.2.3 Adaptive Learning

Markets evolve over time, and a robust POMDP-based trading system must adapt to these changes to remain effective. Adaptive learning algorithms allow the system to update its model parameters in response to new market information Li [2019]. Techniques such as online learning and reinforcement learning enable continuous learning from the market, ensuring that the POMDP model remains attuned to the current market dynamics.

The integration of these scalability and robustness enhancements into POMDP-based trading systems is a testament to the adaptability and potential of POMDPs in the complex domain of financial trading. By addressing the challenges of high-dimensional spaces and the need for robust decision-making, POMDPs stand as a promising avenue for the development of sophisticated trading algorithms that can navigate the intricacies of financial markets with a level of nuance and foresight that traditional models may not provide. The ongoing refinement of these systems, driven by advancements in computational methods and machine learning, holds the promise of unlocking new frontiers in the automation and optimization of trading strategies.

9 Integration with Other Trading Systems

While POMDPs offer a robust framework for decision-making under uncertainty, their integration with existing trading systems can enhance their effectiveness and provide traders with a more comprehensive set of tools. This section discusses the synergies between POMDPs and other trading algorithms, as well as the potential benefits of hybrid approaches that combine POMDPs with other methods.

9.1 Synergies with Existing Trading Algorithms

POMDPs can complement a variety of trading algorithms, including those based on technical analysis, statistical arbitrage, and event-driven strategies. By integrating POMDPs with these algorithms, traders can leverage the strengths of each approach to achieve better performance.

9.1.1 Technical Analysis

Technical analysis involves the study of past market data, primarily price and volume, to forecast future price movements. POMDPs can augment technical analysis by providing a probabilistic framework for decision-making that accounts for the uncertainty inherent in market signals. For instance, a POMDP can be used to determine the optimal time to execute trades based on technical indicators, while considering the likelihood of various market scenarios Murphy [1999].

9.1.2 Statistical Arbitrage

Statistical arbitrage strategies seek to exploit pricing inefficiencies between related financial instruments. POMDPs can enhance these strategies by modeling the uncertainty and dynamics of the spread between pairs of instruments. By incorporating a POMDP, traders can dynamically adjust their positions in response to changes in the estimated probabilities of convergence or divergence in the spread [Gatev et al., 2006].

9.1.3 Event-Driven Strategies

Event-driven strategies are based on trading around corporate events such as earnings announcements, mergers, and acquisitions. POMDPs can be particularly useful in this context by modeling the impact of such events on stock prices and the associated uncertainties. A POMDP framework can help traders assess the risk-reward trade-off of different strategies in the face of event-related information and market reactions [Kritzman et al., 2012].

9.2 Hybrid Approaches Combining POMDPs with Other Methods

Hybrid trading systems that combine POMDPs with other trading methods can benefit from the strengths of each approach. For example, a hybrid system might use machine learning algorithms to predict market trends and a POMDP to manage the execution of trades based on these predictions.

9.2.1 Machine Learning and POMDPs

Machine learning models, such as deep neural networks, have shown promise in predicting market movements based on historical data. However, these models often do not account for the uncertainty in their predictions. By integrating machine learning models with a POMDP framework, traders can incorporate the predictive power of machine learning while managing the uncertainty of the predictions in their decision-making process [Dixon et al., 2018].

9.2.2 Reinforcement Learning and POMDPs

Reinforcement learning (RL) is a type of machine learning where an agent learns to make decisions by interacting with an environment. POMDPs and RL are closely related, as both involve learning policies for decision-making under uncertainty. Combining RL algorithms with POMDPs can lead to more sophisticated trading systems that learn optimal strategies through simulation and real-time market interaction [Moody and Saffell, 2001].

9.2.3 Optimization Techniques and POMDPs

Optimization techniques, such as genetic algorithms and swarm optimization, can be used to fine-tune the parameters of a POMDP model, including the transition and observation probabilities. By integrating these optimization techniques, traders can improve the performance of POMDP-based trading systems by adapting them more effectively to the specific characteristics of the financial markets they are operating in [Brabazon and O'Neill, 2006].

The integration of POMDPs with other trading systems represents a fertile ground for innovation in financial trading. By combining the decision-theoretic rigor of POMDPs with the domain-specific insights provided by other trading algorithms, traders can develop more robust, adaptive, and intelligent trading systems. These hybrid systems have the potential to navigate the complexities of financial markets with a level of sophistication that is greater than the sum of their parts, offering a glimpse into the future of automated trading where human intuition is augmented by advanced computational intelligence.

10 Future Directions and Innovations

The landscape of trading is continuously evolving, driven by advancements in computational techniques and the increasing availability of data. This section explores the emerging trends that are likely to influence the development of POMDP-based trading systems and the potential innovations that could redefine the way trading decisions are made.

10.1 Advancements in Computational Techniques

The computational demands of POMDPs, particularly in high-dimensional spaces, have historically limited their applicability in real-time trading. However, recent advancements in parallel computing, cloud infrastructure, and specialized hardware such as GPUs and TPUs have begun to mitigate these limitations [Raina et al., 2009]. These technologies enable the processing of large-scale data and the execution of complex algorithms at speeds that are becoming compatible with the requirements of high-frequency trading.

10.1.1 Quantum Computing

One of the most anticipated advancements in computational technology is quantum computing. Quantum computers have the potential to perform certain calculations exponentially faster than classical computers. This capability could revolutionize the way POMDPs are solved, allowing for real-time decision-making in environments with a vast number of states and actions [Orús et al., 2019]. Research into quantum algorithms for POMDPs is still in its infancy, but the prospects for their application in trading are promising.

10.2 Machine Learning and Artificial Intelligence

The integration of machine learning and artificial intelligence (AI) with POMDPs is an area of active research. The use of deep learning, for instance, can enhance the state representation and feature extraction processes within POMDP models, leading to more accurate predictions of market dynamics [Deng et al., 2017].

10.2.1 Reinforcement Learning

Reinforcement learning (RL), a subset of machine learning, is particularly relevant to POMDPs due to its focus on learning optimal policies through interaction with an environment. Recent developments in deep reinforcement learning (DRL), which combines deep learning with RL, have shown potential in addressing the curse of dimensionality in POMDPs by efficiently approximating value functions and policies [Mnih et al., 2015]. DRL could enable POMDP-based trading systems to adapt more effectively to changing market conditions and learn complex trading strategies.

10.3 Data-Driven and Adaptive POMDP Models

The proliferation of data in financial markets provides an opportunity to develop more sophisticated POMDP models that are data-driven and adaptive. By leveraging big data analytics, POMDP models can be continuously updated with new information, allowing them to capture the complex and dynamic nature of financial markets more accurately.

10.3.1 Incorporating Alternative Data

Alternative data sources, such as social media sentiment, satellite imagery, and transactional data, can provide additional insights into market movements. Integrating these data sources into POMDP models can enhance the observation process and improve the prediction of market reactions to various events [Bollen et al., 2011].

10.4 Regulatory and Ethical Considerations

As POMDP-based trading systems become more advanced, they will also raise important regulatory and ethical questions. The ability of these systems to rapidly analyze and act on information could lead to concerns about market fairness and transparency. Regulators will need to consider how to oversee the use of AI in trading to ensure that it does not lead to market manipulation or other unfair practices [Challet et al., 2013].

10.5 Conclusion

The future of POMDP-based trading systems lies at the intersection of computational power, machine learning innovation, and the ever-increasing richness of market data. As these elements continue to advance, the potential for POMDPs to transform the trading landscape grows. The integration of quantum computing, deep reinforcement learning, and alternative data into POMDP models promises to create trading systems that are not only more efficient and adaptive but also capable of uncovering complex patterns and strategies that are currently beyond our comprehension. The journey towards this future is fraught with technical challenges and ethical considerations, but the destination holds the promise of a new era in financial decision-making—one that is characterized by a level of sophistication and responsiveness that mirrors the complexity of the markets themselves.

11 Integration with Other Trading Systems

The integration of POMDP-based trading systems with existing trading algorithms and platforms is crucial for their practical deployment in financial markets. This section discusses the potential synergies that can be achieved by combining POMDPs with other trading methods and the challenges that need to be addressed to create effective hybrid trading systems.

11.1 Synergies with Existing Trading Algorithms

POMDP-based trading systems can complement existing trading algorithms by providing a framework for decision-making under uncertainty. For instance, algorithmic trading strategies that rely on technical analysis can benefit from the incorporation of POMDPs by accounting for the probabilistic nature of market movements and the impact of new information [Treleaven et al., 2013].

11.1.1 Combining with High-Frequency Trading

High-frequency trading (HFT) algorithms, which execute trades at very high speeds, can be enhanced by POMDPs to better manage the risks associated with rapid trading in volatile markets. By modeling the uncertainty and partial observability inherent in the market microstructure, POMDPs can help HFT systems make more informed decisions, potentially leading to increased profitability and reduced market impact [Menkveld, 2013].

11.1.2 Risk Management Strategies

Risk management is a critical component of trading. Traditional risk management strategies, such as value-at-risk (VaR) and conditional value-at-risk (CVaR), can be integrated with POMDPs to create a more dynamic approach to managing uncertainty. By considering the sequential nature of trading and the evolution of market conditions, POMDPs can offer a more nuanced view of risk that adapts to new information and market feedback [Jorion, 2007].

11.2 Hybrid Approaches Combining POMDPs with Other Methods

The development of hybrid trading systems that combine POMDPs with other trading methods can leverage the strengths of each approach. For example, POMDPs can be used in conjunction with machine learning models to predict market trends, while the machine learning models can provide the POMDP with a richer set of features for state representation.

11.2.1 Integration with Machine Learning

Machine learning models, particularly those based on time series analysis and natural language processing, can process vast amounts of market data and extract patterns that may not be immediately apparent. When integrated with POMDPs, these models can enhance the observation and state transition functions, leading to more accurate predictions and more effective trading strategies [Buehler et al., 2019].

11.2.2 Reinforcement Learning for Policy Optimization

Reinforcement learning algorithms can be used to optimize the policy of a POMDP-based trading system. By simulating trades within a POMDP framework, RL algorithms can iteratively improve the trading policy to maximize expected returns while controlling for risk. This approach allows for the development of trading strategies that are both robust and adaptive to market changes [Moody and Saffell, 2001].

11.3 Challenges in Integration

While the integration of POMDPs with other trading systems offers many benefits, it also presents several challenges. One of the primary challenges is the alignment of different models and algorithms, which may have been developed with different assumptions and objectives. Additionally, the computational complexity of POMDPs can be a limiting factor when combined with other resource-intensive methods.

11.3.1 Model Alignment and Compatibility

Ensuring compatibility between POMDPs and other trading models requires careful consideration of the interfaces between the systems. This includes the alignment of data formats, synchronization of decision-making cycles, and the reconciliation of potentially conflicting objectives [Nevmyvaka et al., 2006].

11.3.2 Computational Complexity and Resource Allocation

The integration of POMDPs with other trading systems must account for the increased computational load. Efficient resource allocation and the use of advanced computational techniques are necessary to maintain the responsiveness required for real-time trading [Silver et al., 2016].

The successful integration of POMDP-based trading systems with existing algorithms and platforms has the potential to create a new class of trading systems that are both intelligent and adaptable. These hybrid systems could offer a significant competitive advantage by effectively navigating the complexities of financial markets. As the field progresses, the fusion of POMDPs with diverse trading methodologies will likely lead to innovative solutions that redefine the boundaries of algorithmic trading.

12 Future Directions and Innovations

The landscape of algorithmic trading is continuously evolving, driven by advancements in computational techniques and the increasing availability of data. This section explores the emerging trends that are likely to influence the development of POMDP-based trading systems and the potential innovations that could enhance their performance and applicability in the financial markets.

12.1 Advancements in Computational Techniques

The computational demands of POMDPs, particularly in high-dimensional spaces, have historically limited their practical application in trading. However, recent advancements in computational power and algorithms offer new opportunities to overcome these challenges.

12.1.1 Quantum Computing

Quantum computing presents a promising avenue for solving complex POMDPs that are intractable with classical computers. The quantum parallelism inherent in quantum computers could potentially allow for the simultaneous evaluation of multiple market scenarios, leading to more efficient policy optimization and state estimation in POMDPs [Orús et al., 2019].

12.1.2 Distributed and Parallel Computing

The use of distributed and parallel computing frameworks can significantly reduce the time required to solve POMDPs by leveraging multiple processors to perform computations concurrently. Cloud computing platforms and specialized hardware, such as graphics processing units (GPUs), are becoming more accessible and can facilitate the scaling of POMDP-based trading systems to handle larger and more complex models [Dean et al., 2012].

12.2 Machine Learning and Artificial Intelligence

The integration of machine learning and artificial intelligence (AI) with POMDPs is a rapidly growing area of research. These technologies can enhance various aspects of POMDP-based trading systems, from data preprocessing to policy learning.

12.2.1 Deep Learning for State Representation

Deep learning techniques, particularly deep neural networks, can be employed to learn complex representations of market states from raw financial data. These representations can capture nonlinear relationships and latent factors that are not easily modeled with traditional approaches, potentially leading to more accurate predictions and better trading performance [Goodfellow et al., 2016].

12.2.2 Reinforcement Learning Enhancements

Reinforcement learning (RL) algorithms, especially those incorporating deep learning (deep RL), have shown promise in learning optimal policies for POMDPs. Innovations in RL, such as policy gradient methods and actor-critic architectures, can be adapted to the trading domain to refine the decision-making process of POMDP-based trading systems [Mnih et al., 2015].

12.3 Data-Driven and Adaptive Systems

The dynamic nature of financial markets necessitates trading systems that can adapt to changing conditions. POMDP-based trading systems must be capable of learning from new data and adjusting their strategies accordingly.

12.3.1 Online Learning and Adaptation

Online learning algorithms enable POMDP-based trading systems to update their models in real-time as new market data becomes available. This continuous learning process allows the systems to remain relevant and effective in the face of market volatility and evolving trends [Hoi et al., 2018].

12.3.2 Transfer Learning for Market Shifts

Transfer learning techniques can be utilized to adapt pre-trained POMDP models to new market conditions or asset classes with minimal retraining. By leveraging knowledge from related domains, transfer learning can expedite the adaptation process and improve the generalizability of trading strategies [Pan and Yang, 2010].

12.4 Regulatory Compliance and Ethical Considerations

As POMDP-based trading systems become more sophisticated, it is essential to consider the regulatory and ethical implications of their deployment. Ensuring compliance with financial regulations and maintaining transparency in algorithmic decision-making are critical for the sustainable integration of these systems into the markets.

12.4.1 Explainable AI for Trading Decisions

The development of explainable AI (XAI) methods can help demystify the decisions made by POMDP-based trading systems. By providing interpretable justifications for trading actions, XAI can facilitate regulatory compliance and build trust among market participants [Arrieta et al., 2020].

12.4.2 Mitigating Market Manipulation Risks

The potential for algorithmic trading systems to inadvertently engage in market manipulation or exacerbate financial crises must be addressed. Robust design and testing of POMDP-based trading systems, along with appropriate safeguards, are necessary to prevent such outcomes and ensure market stability [Kirilenko and Lo, 2017].

The future of POMDP-based trading systems lies at the intersection of technological innovation and responsible deployment. By embracing the advancements in computational techniques and machine learning, while remaining cognizant of regulatory and ethical considerations, these systems can lead to a new era of intelligent and adaptive trading. The journey ahead is not without challenges, but the potential rewards for the financial industry and the broader economy are substantial, promising a future where the enigmatic nature of markets can be navigated with unprecedented precision and insight.

References

Román Orús, Samuel Mugel, and Enrique Lizaso. Quantum computing for finance: Overview and prospects. *Reviews in Physics*, 4:100028, 2019.

William F. Sharpe. The Sharpe Ratio. *The Journal of Portfolio Management*, 21(1):49–58, 1994.

Anthony Brabazon and Michael O'Neill. Biologically Inspired Algorithms for Financial Modelling. *Springer*, 2006.

Albert J. Menkveld. High-frequency trading and the new market makers. *Journal of Financial Markets*, 16(4):712–740, 2013.

Mark Kritzman, Yuanzhen Li, Sebastien Page, and Roberto Rigobon. Principal Components as a Measure of Systemic Risk. *Journal of Portfolio Management*, 38(4):112–126, 2012.

Yi Shen, Jinjie Zhang, and Xun Yu Zhou.
ewblock Adaptive trading with a Markov decision process model.
ewblock *Operations Research*, 62(5): 998–1012, 2014.

Álvaro Cartea, Sebastian Jaimungal, and José Penalva. Algorithmic and High-Frequency Trading. *Mathematics, Finance and Risk*, pages 1–30. Cambridge University Press, 2015.

Christopher J.C.H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3-4):279–292, 1992.

Robert Almgren and Neil Chriss. Optimal execution of portfolio transactions. *Journal of Risk*, 3:5–39, 2001.

Richard S. Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. *In Advances in Neural Information Processing Systems*, pages 1057–1063, 2000.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Deep Learning. *MIT Press*, 2016.

Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, Raja Chatila, and Francisco Herrera. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58:82–115, 2020.

John Moody and Matthew Saffell. Learning to Trade via Direct Reinforcement. *IEEE Transactions on Neural Networks*, 12(4):875–889, 2001.

Andrei Kirilenko and Andrew W. Lo. Flash Crash: The Impact of High Frequency Trading on an Electronic Market. In *The Journal of Finance*, 72(3):967–998, 2017.

Jeffrey Dean, Greg Corrado, Rajat Monga, Kai Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao, Marc'Aurelio Ranzato, Andrew Senior, Paul Tucker, Ke Yang, and Andrew Y. Ng. Large scale distributed deep networks. In *Advances in Neural Information Processing Systems*, pages 1223–1231, 2012.

Marcos López de Prado. The sequential information bottleneck for financial modeling. *Quantitative Finance*, 12(4):577–585, 2012.

James D. Hamilton. Analysis of Time Series Subject to Changes in Regime. *Journal of Econometrics*, 45(1-2):39–70, 1990.

Lawrence R. Rabiner. An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*, 3(1):4–16, 1986.

Young Bin Kim. Predicting Stock Market Fluctuations Using LDA Topic Modeling and Machine Learning. *Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC)*, pages 120–121, 2014.

John Moody and Matthew Saffell. Learning to trade via direct reinforcement. *IEEE Transactions on Neural Networks*, 12(4):875–889, 2001.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

Bradley Efron and Robert J. Tibshirani. *An Introduction to the Bootstrap*. CRC press, 1994.

Sinno Jialin Pan and Qiang Yang. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.

Rajat Raina, Anand Madhavan, and Alexander Colville. Large-scale automatic classification of phishing pages. In *Advances in Neural Information Processing Systems*, pages 545–552, 2009.

Myles Hollander, Douglas A. Wolfe, and Eric Chicken. *Nonparametric Statistical Methods*. John Wiley & Sons, 3rd edition, 2013.

Paul C. Tetlock. Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):1139–1168, 2007.

Ruey S. Tsay. Analysis of Financial Time Series. *Wiley-Interscience*, 2005.

John J. Murphy. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. New York Institute of Finance, 1999.

Philip Treleaven, Michal Galas, and Vikram Lal. Algorithmic trading review. *Communications of the ACM*, 56(11):76–85, 2013.

Joel Hasbrouck. Empirical market microstructure: The institutions, economics, and econometrics of securities trading. *Oxford University Press*, 2007.

Terry W. Young. Calmar Ratio: A Smoother Tool. *Futures*, 20(1):40, 1991.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition, 2018.

Johan Bollen, Huina Mao, and Xiaojun Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8, 2011.

Matthew Gentzkow and Jesse M. Shapiro.
ewblock Polarization in 20th century America.
ewblock In *Handbook of Economic Growth*, volume 2, pages 3–69. Elsevier, 2016.

William F. Sharpe. The Sharpe Ratio. *The Journal of Portfolio Management*, 21(1):49–58, 1994.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Damien Challet, Ahmed Bel Hadj Ayed, and Yannick Lempérière. Market bugs: Long-range dependence and market making. In *Econophysics of Agent-Based Models*, pages 25–48. Springer, 2013.

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

John J. Murphy. Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. *New York Institute of Finance*, 1999.

Leslie P. Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, 1998.

Philippe Jorion. Value at Risk: The New Benchmark for Managing Financial Risk. *McGraw-Hill*, 1997.

Richard Bellman. A Markovian Decision Process. *Journal of Mathematics and Mechanics*, 6(5):679–684, 1957.

Ananth Madhavan. Market Microstructure: A Survey. *Journal of Financial Markets*, 3(3):205–258, 2002.

Robert Almgren and Neil Chriss. Optimal Execution of Portfolio Transactions. *Journal of Risk*, 3:5–39, 2001.

Steven C. H. Hoi, Doyen Sahoo, Jing Lu, and Peilin Zhao. Online Learning: A Comprehensive Survey. *arXiv preprint arXiv:1802.02871*, 2018.

Peng Zhang and Yanchun Zhang. The application of genetic algorithms in stock market data mining optimisation. In *Advances in Natural Computation*, pages 823–831, Springer, 2005.

Malik Magdon-Ismail, Amir F. Atiya, Amrit Pratap, and Yaser S. Abu-Mostafa. On the Maximum Drawdown of a Brownian Motion. *Journal of Applied Probability*, 41(1):147–161, 2004.

Ronald A. Howard. Risk-Sensitive Markov Decision Processes. *Management Science*, 18(7):356–369, 1972.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

Nicole Bäuerle and Ulrich Rieder. *Markov Decision Processes with Applications to Finance*. Springer, 2011.

Evan Gatev, William N. Goetzmann, and K. Geert Rouwenhorst. Pairs Trading: Performance of a Relative-Value Arbitrage Rule. *Review of Financial Studies*, 19(3):797–827, 2006.

Ruey S. Tsay. Analysis of Financial Time Series. *Wiley Series in Probability and Statistics*, pages 1–25. Wiley-Interscience, 2005.

Arnaud Doucet, Nando de Freitas, and Neil Gordon. Sequential Monte Carlo Methods in Practice. *Springer*, 2000.

Siddharth G. Patro and Bhabani S. Sahu. Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*, 2015.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin.
ewblock Attention is all you need.
ewblock *In Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

Guy Shani, Joelle Pineau, and Robert Kaplow. A survey of point-based POMDP solvers. *Autonomous Agents and Multi-Agent Systems*, 27(1):1–51, 2013.

Richard S. Sutton and Andrew G. Barto. Reinforcement Learning: An Introduction. *MIT Press*, 2nd edition, 2018.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement Learning for Optimized Trade Execution. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680, 2006.

Eugene F. Fama and Kenneth R. French. Permanent and Temporary Components of Stock Prices. *Journal of Political Economy*, 96(2):246–273, 1988.

Cameron Browne, Edward Powley, Daniel Whitehouse, Simon Lucas, Peter I. Cowling, Philipp Rohlfshagen, Stephen Tavener, Diego Perez, Spyridon Samothrakis, and Simon Colton. A Survey of Monte Carlo Tree Search Methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1):1–43, 2012.

Ronald A. Howard and James E. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, 1972.

Hans Buehler, Lukas Gonon, Josef Teichmann, and Ben Wood. Deep hedging. *Quantitative Finance*, 19(8):1271–1291, 2019.

Stephen J. Brown and Robert H. Jennings. On technical analysis. *Review of Financial Studies*, 2(4):527–551, 1989.

Frank A. Sortino and Lee N. Price. Performance measurement in a downside risk framework. *The Journal of Investing*, 3(3):59–64, 1994.

Yuriy Nevmyvaka, Yi Feng, and Michael Kearns. Reinforcement learning for optimized trade execution. In *Proceedings of the 23rd International Conference on Machine Learning*, pages 673–680, 2006.

Yuxi Li. Deep Reinforcement Learning: An Overview. *arXiv preprint arXiv:1701.07274*, 2019.

Peter J. Brockwell and Richard A. Davis. Introduction to Time Series and Forecasting. *Springer*, 2nd edition, 2002.

Christian T. Brownlees and Giampiero M. Gallo. Comparison of volatility measures: a risk management perspective. *Journal of Financial Econometrics*, 8(1):29–56, 2010.

George E. Monahan. A survey of partially observable Markov decision processes: Theory, models, and algorithms. *Management Science*, 28(1):1–16, 1982.

Matthew Dixon, Diego Klabjan, and Jin Hoon Bang. Sequence Classification of the Limit Order Book Using Recurrent Neural Networks. *Journal of Computational Science*, 24:277–286, 2018.

Xin Huang, Heping Pan, and Yanhui Chen.
ewblock Portfolio selection with a new definition of risk.
ewblock *European Journal of Operational Research*, 207(2): 851–865, 2010.

Maureen O'Hara. Market Microstructure Theory. *Blackwell Publishers*, 1995.

Satinder P. Singh and Richard S. Sutton. Reinforcement Learning with Replacing Eligibility Traces. *Machine Learning*, 22(1-3):123–158, 1995.

Román Orús, Samuel Mugel, and Enrique Lizaso. Quantum computing for finance: Overview and prospects. *Reviews in Physics*, 4:100028, 2019.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *In Proceedings of the 34th International Conference on Machine Learning*, pages 1126–1135, 2017.

Harry Markowitz. Portfolio Selection. *The Journal of Finance*, 7(1):77–91, 1952.

James D. Hamilton. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2):357–384, 1989.

Michael Kearns, Yuriy Nevmyvaka, and Jennifer Wortman Vaughan.
ewblock Machine learning for market microstructure and high-frequency trading.
ewblock *High-Frequency Trading: New Realities for Traders, Markets, and Regulators*, pages 91–113. Risk Books, 2013.

Mohammad Ghavamzadeh, Shie Mannor, Joelle Pineau, and Aviv Tamar. Bayesian Reinforcement Learning: A Survey. *Foundations and Trends in Machine Learning*, 8(5-6):359–483, 2015.

Ian T. Jolliffe. Principal component analysis. *Springer*, 2016.

M. Kabir Hassan and B. Nath. Stock Market Forecasting Using Hidden Markov Model: A New Approach. In *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)*, pages 192–196, 2005.

Stéphane Ross, Joelle Pineau, Sébastien Paquet, and Brahim Chaib-draa. Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*, 32:663–704, 2008.

Joelle Pineau, Geoffrey Gordon, and Sebastian Thrun. Point-based value iteration: An anytime algorithm for POMDPs. *In Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 1025–1032, 2003.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, and others. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

Philippe Jorion. Value at risk: The new benchmark for managing financial risk. *McGraw-Hill*, 3rd edition, 2007.

Yue Deng, Feng Bao, Youyong Kong, Zhiquan Ren, and Qionghai Dai. Deep direct reinforcement learning for financial signal representation and trading. *IEEE transactions on neural networks and learning systems*, 28(3):653–664, 2017.

Peter J. Brockwell and Richard A. Davis. Introduction to Time Series and Forecasting. *Springer Texts in Statistics*, pages 1–19. Springer, 2002.

John J. Murphy. Technical analysis of the financial markets: A comprehensive guide to trading methods and applications. *New York Institute of Finance*, 1999.