UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO

PROJETO DE PESQUISA FAPESP

Aprendizado por Reforço para otimização de uma estratégia de Market-Making

LINHA DE FOMENTO: BOLSA NO PAÍS - REGULAR - INICIAÇÃO CIENTÍFICA

Candidato:

Rafael Zimmer

Orientador:

São Paulo

October 24, 2023

Resumo

Neste projeto de pesquisa propomos realizar uma análise de estratégias financeiras existentes para criação de mercados (*market-making* em inglês), assim como o possível uso de técnicas de Aprendizado por Reforço (*AR*) para maximização do retorno esperado de tais estratégias. Recentes avanços literários na área de *AR* para otimização de agentes financeiros focam na busca por políticas que maximizem o retorno diário e minimizem o risco das ordens e do inventário gerenciado por tais agentes. O risco associado a uma estratégia de *market-making* vem da falta de garantia para execução das ordens criadas, e dependendo do processo de chegada das ofertas de compra e venda, pode ser que o agente não tenha nenhuma de suas ordens executadas, ou até mesmo feche o dia com um retorno negativo. Considerando essa definição de risco e uma lacuna na literatura voltada à minimização do risco pós-fechamento de mercado — chamado de risco *overnight* — realizaremos uma pesquisa bibliográfica sobre técnicas de *AR* para otimização de políticas de negociação, assim como a conceitualização e treino de um agente de *MM* que maximize o retorno diário esperado sob a restrição de finalizar o dia sem posição remanescente, ou seja, zerar o risco noturno associado.

Contents

1 Introdução e justificativa

Há um crescente foco na literatura da computação financeira para com estratégias intra-diárias de *market-making* (Guéant, 2017), que consistem em criar ordens de limite buscando lucrar em cima da diferença entre o preço de bid (Melhor preço de compra) e o preço de ask (Melhor preço de venda), chamada de Bid-Ask-Spread (*BAS*). Por serem estratégias intra-diárias, utilizam técnicas de gerenciamento de risco estritamente voltadas para os horários em que o mercado está aberto e permite a criação de novas ordens. Durante esses horários, um agente de *market-making* (*MM*) pode ofertar simultaneamente ordens limite de compra e venda, sob o risco de poder não ter uma ordem executada até o período de fechamento do mercado (Guéant et al., 2012) (Selser et al., 2021) (Bakshaev, 2020).

Se o agente não tem suas ordens finalizadas, ou escolhe manter um inventário após o fechamento do pregão, estará se sujeitando ao risco associado à posição durante a noite, chamado também de risco noturno, ou *overnight*. Esse tipo de risco advém do fato de que o mercado não processa nenhuma ordem existente durante a noite até sua abertura no próximo dia. Logo, qualquer posição *overnight* está vulnerável a eventos inesperados durante esse período, como notícias financeiras, mudanças macroeconômicas e outros que possam resultar em aberturas de mercado voláteis potencialmente desfavoráveis à posição mantida. Um agente capaz de **maximizar o retorno das operações diárias** e **minimizar o risco associado à posição noturna** seria portanto uma contribuição crítica para estratégias de *MM* existentes usadas por fundos de investimento, corretoras e bancos.

Com o objetivo de obter uma política de criação de ofertas condicionada a minimizar o risco *overnight* teremos como primeiro desafio determinar uma metodologia para obter dados financeiros e validar nossas hipóteses sobre o ambiente de simulação. É necessário ter em mente que as ordens criadas geram mudanças no estado do mercado e conseguintemente alteram a amostra histórica. Uma amostragem não estática impede o uso de técnicas para *backtesting* regulares, pois um agente de MM não é um consumidor de preços (*price-taker* em inglês), mas sim um fornecedor de preços (*price-maker*). Em um cenário real, não é ideal desconsiderar o impacto das interações do agente com o mercado: um agente de *MM* que esteja atuando no mercado tentará sempre manter suas ofertas no melhor patamar de preço possível, de modo a ter suas ordens executadas primeiro. Assim, o envio de novas ofertas afetará as ordens subsequentes deste outro agente.

Outro detalhe importante sobre o ambiente é a frequência dos processos de chegada das ordens. Uma nova oferta que ultrapasse o melhor preço do disponível no mercado causa quase de imediato

(a depender da corretora) a atualização do livro de ordens. O tempo de chegada irá afetar muito mais a política da estratégia ótima obtida quando comparado com dados usados para estratégias de longo prazo ou de *price-taking*.

Em suma, tem-se duas etapas adicionais a serem consideradas em conjunto do problema principal de conceitualizar e treinar um agente:

- uso de dados históricos estáticos para simulação de um ambiente dinâmico;
- tratamento de dados com alta frequência que possam influenciar a política obtida;

Como conclusão da pesquisa, esperamos obter um agente que execute uma estratégia de *MM* que maximize o valor esperado do retorno diário em múltiplos ativos e que se adapte aos processos estocásticos de chegada de ofertas e transações. Iremos impor a restrição de um risco noturno máximo associado ao inventário do agente.

2 Das áreas a serem abordadas

2.1 Market Making

O market-making é uma forma de negociação (*trading* em inglês) que de forma simplificada consiste em comprar certo ativo a um determinado preço e vender-lo a um preço maior. Como os preços de compra e venda dos ativos são afetados pela demanda e oferta no momento, cada agente no mercado exerce certa influência sobre o valor de um ativo, a depender das ofertas que o mesmo mantém para tal ativo no livro de ordens limite (*limit order book*, ou *LOB*).

No *LOB* os registros se dão primeiramente por ordem de preço, e em segundo por data de criação: as ofertas com melhor preço (tanto para compra como para venda) ficam no topo do livro, e caso tenham o mesmo valor entre si tem sua posição desempatada pela ordem temporal de chegada. No livro de compras, a melhor oferta é a cuja ordem oferece o maior preço, e o contrário vale para o livro de vendas.

Nas bolsas de valores digitais a execução de uma transação é automática e auxiliada por um sistema chamado de *matching engine* - ou seja, um motor para pareamento de ordens. Esse sistema verifica se para a melhor oferta de compra (ou venda) existe outra correspondente no livro de venda

(ou compra) com valor menor ou igual (ou maior ou igual para venda). Após o pareamento, a bolsa anuncia o a execução da ordem e as ofertas relacionadas são removidas dos livros.

Em suma, os principais elementos do market making incluem, mas não se limitam à:

Spread: É a diferença entre o preço de compra (*bid* em inglês) e o preço de venda (*ask* em inglês) entre duas ofertas. O market maker busca lucrar com a diferença entre esses preços.

Livro de Ordens Limite: É o conjunto ordenado onde as ofertas de compra e venda são registradas. Também é permitido o ajuste dos preços de compra e venda de ofertas existentes por parte dos agentes.

Gestão de Risco: Todos *market-makers* enfrentam riscos em suas negociações, entre eles o risco de inventário e risco de mercado. O risco de inventário ocorre quando o market maker mantém uma posição desequilibrada entre ativos comprados e vendidos, enquanto o risco de mercado está relacionado às flutuações nos preços dos ativos.

No contexto deste projeto, o foco da pesquisa será a otimização de uma estratégia de *market-making* que minimize o risco de inventário durante a noite. A estratégia será composta por um agente responsável pela interação com o mercado e alocação de preços sob políticas para redução de risco. O Aprendizado por Reforço (RL) foi a abordagem escolhida para obter um agente ótimo capaz de realizar essa tarefa.

2.2 Aprendizado por Reforço

O Aprendizado por Reforço (RL) é um paradigma de aprendizado de máquina baseado em princípios da psicologia comportamental e em otimização estocástica, especificamente tarefas de controle ótimo. Simplificadamente, trata-se de uma técnica para modelagem, simulação e treino de um agente capaz de interagir com um ambiente dinâmico de modo a maximizar uma recompensa cumulativa ao longo do tempo.

O processo de RL é análogo ao modo como os seres humanos aprendem por tentativa e erro. Um agente explora diferentes ações, observa as consequências dessas ações no ambiente e ajusta sua estratégia com base nas recompensas obtidas e no novo estado do ambiente. O objetivo final é a obtenção de uma política que maximize a recompensa esperada.

Essencialmente, um agente consiste de três componentes:

Política (ou *Policy*): Define o processo de tomada de decisões do agente, ou seja, como ele escolhe ações em resposta às observações do ambiente. Pode ser uma estratégia determinística ou estocástica.

Recompensa (ou *Reward*): É uma medida numérica que permite ao agente interpretar quão boa ou ruim foi uma ação específica em função do estado atual do ambiente. O objetivo do agente é maximizar a recompensa cumulativa ao longo do tempo.

Modelo do Ambiente (ou *Environment*): Representa e simula o estado do ambiente real, assim como as possíveis ações que o agente possa realizar. É uma descrição quantitativa de elementos do ambiente e como as ações tomadas pela política afetam o ambiente em si.

O RL pode ser aplicado à diversos domínios, da robótica, jogos à finanças e no contexto deste projeto será o paradigma central na idealização e modelagem da estratégia de *market-making*.

3 Compreendendo o Problema Inicial

3.1 Definição formal do *market making*

Os mercados normalmente definem uma quantidade mínima por lote ofertado. Se uma bolsa determina um mínimo de 100 ações um vendedor não poderia criar uma oferta de 50 ações. Se os preços ofertados no momento permitirem a execução de um negócio, a transação ocorrerá em cima da menor quantidade entre as duas ofertas, ou seja, se uma ordem de venda de 100 ações fecha um negócio com uma ordem de compra de 200 ações por exemplo, a quantidade negociada será de apenas 100. O restante da oferta de compra, i.e. 100 ações, continua no livro de oferta de compras, enquanto no lado da venda, a próxima oferta de maior preferência vai para o topo do livro.

Definimos como p^a o preço de uma determinada oferta de venda, chamado de *ask* em inglês e p^b para compras, ou *bid* em inglês. Um agente pode ter diversas ofertas para cada ativo i, em tempos diferentes indexados pelo símbolo t:

$$p^a_{t,i} \text{ para ofertas de venda e}$$

$$p^b_{t,i} \text{ para ofertas de compra.}$$

A posição de uma oferta no livro de vendas para um ativo i é dada pela função $\mathbf{pos}^a_{t,i}(p) = k$ está na posição k da fila [1]. A diferença entre ofertas de compra e venda para um mesmo ativo $\Delta_{t,i} = p^a_{t,i} - p^b_{t,i}$ é o *spread* no momento t e caso $\mathbf{pos}^a_{t,i}(p) = \mathbf{pos}^b_{t,i}(p) = 0$ é chamado de *bid-ask spread* $\boldsymbol{\Delta}_{t,i}$.

[1] onde $k = 0$ é a melhor oferta; com valores de preço decrescentes para vendas, e crescentes para compras.

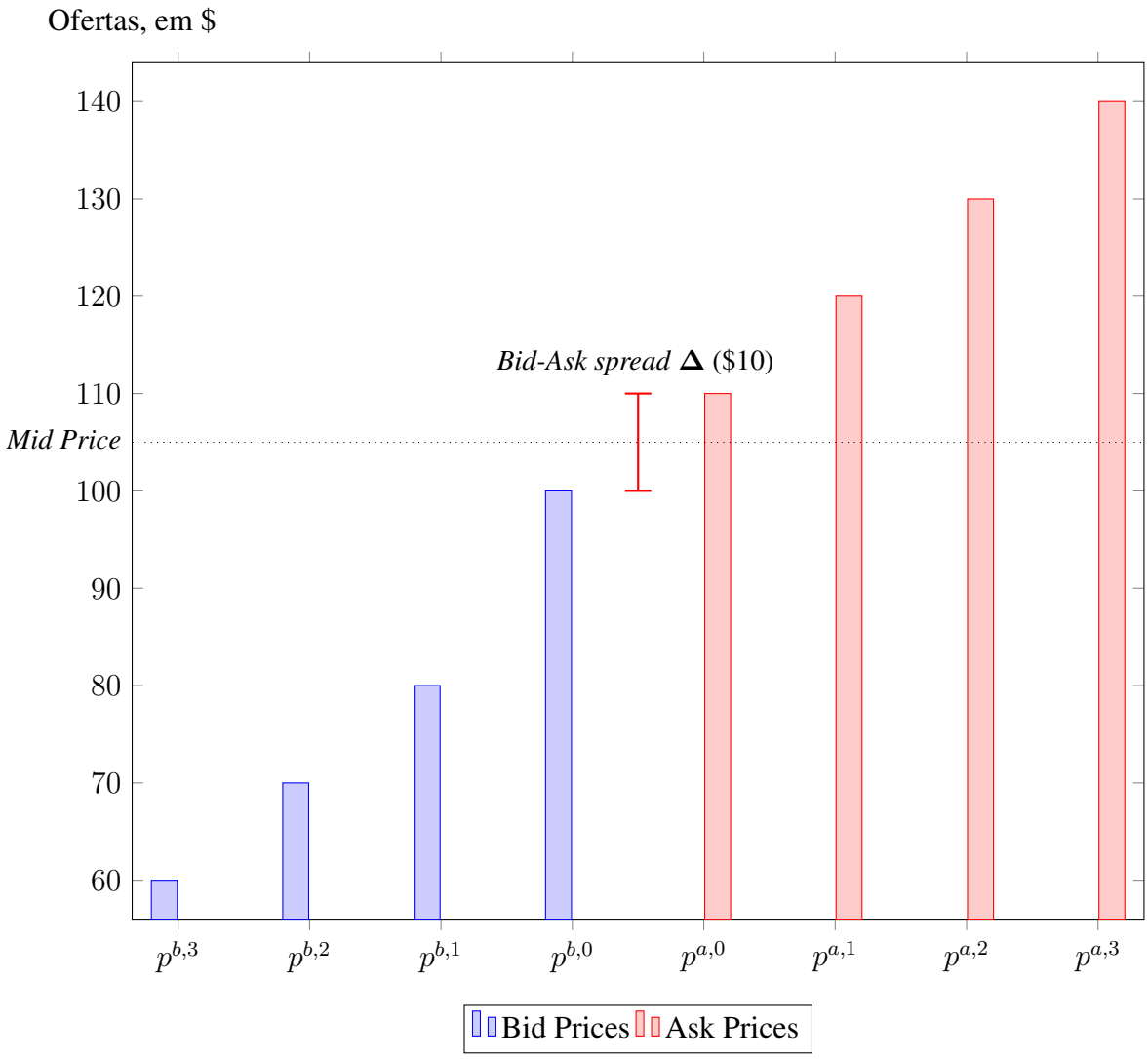

Figure 1: Gráfico de ofertas de um livro de ordens limite L qualquer

O objetivo de uma estratégia de *market making* (*MM*) nesse contexto é criar ofertas de compra com valor maior que a melhor oferta, ou menor para ofertas de venda que venham a se tornar as melhores ofertas:

- de venda, tais que $p^a < P^a$; ou
- de compra, tais que $p^b > P^{b,}$;

onde p é o preço da oferta do agente e P é a melhor oferta existente $\mathbf{pos}^a_{t,i}(P) = \mathbf{pos}^b_{t,i}(P) = 0$. É importante notar que quaisquer ordens criadas por um agente de *MM* não geram novas transações no instante *t* (indicado por $p^a > P^b$ e $p^b < P^a$ para um mesmo ativo). A seguir um exemplo da reta de preços para um momento t e um ativo i qualquer, onde $p^{b,k}$ é tal que $\mathbf{pos}^b(p^{b,k}) = k$:

Para ilustrar melhor o funcionamento do livro de ordens considere a seguinte situação para um

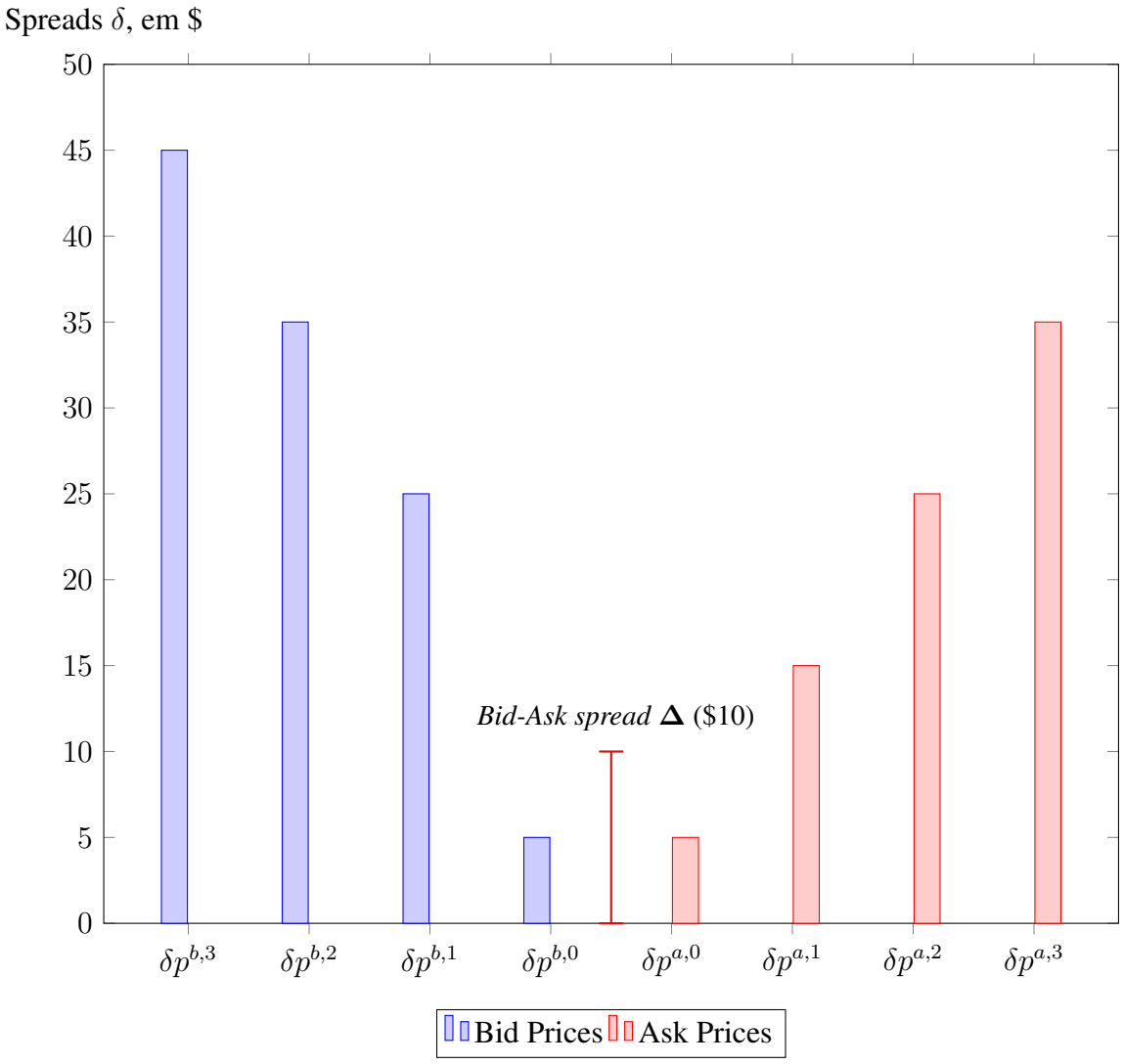

Figure 2: Gráfico de spreads de um livro de ordens limite L qualquer

mesmo momento e ativo: um agente qualquer de *MM* tem a **única oferta de venda** pelo preço de p^a no mercado. Caso surja uma nova oferta de **compra** com melhor preço e acima do preço da oferta do agente $P^b \geq p^a$, será gerada uma transação pelo preço p^a e a ordem de preço P^b que gerou a transação é chamada de ordem de mercado, enquanto a ordem do agente é chamada de ordem limite. Após o motor de transações da bolsa receber a ordem com preço P, uma transação ocorre e ambas ofertas são removidas do livro de ordens. Em seguida, o agente tem sua posição r_t ajustada:

$$\begin{aligned} r_{t+1} &:= r_t + p^a q \text{ para vendas e} \\ r_{t+1} &:= r_t - p^b q \text{ para compras} \end{aligned} \tag{1}$$

sendo r_t o valor da posição do agente no instante t e q a quantidade de ativos negociadas na transação em questão.

O agente em seguida escolhe esperar ou realizar uma das ações abaixo:

1. inserir uma nova oferta de venda, substituindo a oferta anterior;
2. inserir ou ajustar uma oferta de compra existente no livro de ofertas de compras;

A decisão do agente irá depender de sua expectativa sobre a evolução do mercado, assim como do processo de chegada de ordens de outros agentes, representados pelo identificador n, incluindo, mas não limitado à

- probabilidade de chegar uma oferta de compra com preço superior $Pr(P^b_{t+s,i} > p^b_{t,i})$, em algum momento no futuro $s > 0$;
- probabilidade de chegar uma oferta de venda com preço inferior $Pr(P^a_{t+s,i} < p^a_{t,i})$, em algum momento no futuro $s > 0$;
- liquidez esperada para o mercado a partir de t até o momento de fechamento do pregão T;
- risco futuro da posição ultrapassar os limites estabelecidos pelas corretoras.

De modo a simplificar as equações adiantes e utilizar uma notação mais comumente usada no contexto de ordens limite, definimos:

$\delta_t(p)$**:** $\mathbb{R} \to \mathbb{R} = |p - p_t|$ como a diferença entre o preço p de uma oferta de venda ou compra e o preço de mercado p da ação subjacente no momento t.

$\delta_t(p, q)$**:** $\mathbb{R}^2 \to \mathbb{R} = q \cdot \delta_t(p)$ como a função que mapeia o impacto de q ações negociadas na carteira do agente sob o *spread* parcial δ_t.

r_t é o retorno obtido pelo agente no momento t. É calculado pela soma do impacto de todos ativos, onde $o_t = \{(p^a_{t,0}, Q^a_{t,0}), ..., (p^a_{t,n}, Q^a_{t,n}), (p^b_{t,0}, Q^b_{t,0}), ..., (p^b_{t,m}, Q^b_{t,m})\}$ é o conjunto de ofertas do agente para todos ativos no momento t. Cada oferta é uma tupla (p, Q) de preço por ativo e quantidade de ativos ofertada.

$$\begin{aligned} r_t = &\sum_{i=0}^{n} \delta_t(p^a_{t,i}, q^a_{t,i}) \\ &- \sum_{i=0}^{m} \delta_t(p^b_{t,i}, q^b_{t,i}) \\ &\forall t < T \end{aligned} \tag{2}$$

Onde os valores de q são as quantidades efetivamente executadas da ordem, podendo ser menor (no caso de uma ordem parcial) ou igual à Q (ordem total), onde Q é a quantidade inicialmente ofertada pelo agente. Para considerar custos de transação — que incluem tipicamente custos da corretora e emolumentos das bolsas — basta alterar a expressão para $r_t := r_t - c$, sendo c o custo total de todas transações realizadas.

O agente de *MM*, por fim, observa a sua posição acumulada durante todo o período de negociação T para decidir se obteve retorno positivo ou negativo:

$$R_T = \sum_{t=0}^{T} r_t \tag{3}$$

Define-se matematicamente o valor da posição R_T como a agregação das receitas de vendas, e dos custos de compra e de transações até o momento T.

3.2 Modelagem do agente de *market making* e do objetivo

O objetivo principal do agente de *MM* é decidir dentro do intervalo de preços possíveis para uma ação o valor que proporcione o maior retorno para o menor risco associado, de acordo com a fronteira do mercado eficiente (Markowitz, 1952). O agente também pode decidir a quantidade de ações ofertadas por determinado preço, mas não tem controle direto sobre as quantidades efetivamente negociadas. Ou seja, a quantidade executada q é uma variável estocástica, tal que $P(q = Q),\ \forall q <=$

Q é a probabilidade de que uma oferta de Q ações seja executada por completo em uma ordem.

Podemos consequentemente modelar o agente como um problema de otimização estocástica com restrições (também chamado de programação estocástica), onde o objetivo inicial do agente é separado em duas etapas:

1. maximização do *bid-ask spread* $\mathbf{\Delta}_{t,i} = \delta_{t,i}(p^a) + \delta_{t,i}(p^b) = |p^a - p^b|$ para todos ativos;

2. maximização da quantidade executada esperada $\mathbb{E}[q^a_{t,i}]$ e $\mathbb{E}[q^b_{t,i}]$ de ordens de venda e compra realizadas em cima do maior *bid-ask spread* $\mathbf{\Delta}_{t,i}$, garantindo que a negociação ocorra na fronteira eficiente.

As variáveis de decisão são as combinações possíveis de ofertas de venda e compra — ou seja, combinações do conjunto de ofertas do agente o_t[2]. A função objetivo do problema é o valor esperado do retorno diário, considerando a incerteza da quantidade executada $q_{t,i} \leq Q_{t,i}$ por ordem. Substituindo o valor de q na equação 2 pelo seu valor esperado:

$$\begin{aligned} \mathbb{E}[r_t] = &\sum_{i=0}^{n} \delta_t(\, p^a_{t,i}\,,\ \mathbb{E}[\, q^a_{t,i}\,]) \\ &- \sum_{i=0}^{m} \delta_t(\, p^b_{t,i}\,,\ \mathbb{E}[\, q^b_{t,i}\,]) \\ &\qquad\qquad \forall t < T \end{aligned} \tag{4}$$

e o retorno diário acumulado da equação 3 é utilizado como função objetiva que se deseja maximizar:

$$\max_{A,B} \quad \sum_{t=0}^{T} \mathbb{E}[r_t] \tag{5}$$

De modo a obter uma solução para a otimização estocástica definimos o agente como um processo de decisão de Markov $(\mathcal{S}, \mathcal{A}, T, r)$, buscando modelar o problema para o paradigma de Aprendizado por Reforço:

$\mathcal{S}$ é o espaço de estados possíveis, representado pelo conjunto $\{o_t, L_t) \mid t < T\}$, onde cada estado $s \in \mathcal{S}$ é uma combinação possível de ofertas do agente o, e o livro de ordens limite L no momento t;

[2] Note que o agente não decide a quantidade executada q, apenas a quantidade ofertada Q

$\mathcal{A}$ é o espaço de ações que o agente pode realizar, ou seja, a combinação de novos *spreads* $\delta(p_{t+1})$ e novas quantidades Q_{t+1} para cada tupla (p, Q) do conjunto o_t de ofertas de venda e compra existentes;

T são as transições possíveis entre estados dado uma ação tomada pelo agente. São representadas pela função de transição $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$, que mapeia o estado atual e a ação tomada para a probabilidade de ir para um estado subjacente. No caso do agente de *market-making*, a função T recebe o estado atual s e a ação a tomada pelo agente (conjunto de *spreads* e quantidades ofertadas atualizadas). Em seguida recebe um possível estado futuro $s' \in \mathcal{S}$ e retorna a probabilidade de transição $T(s, a, s') = Pr(S_{t+1} = s' \mid S_t = s, A = a)$;

r é a função de recompensa da cadeia aleatória, que mapeia o estado atual e a ação do agente para a probabilidade de uma recompensa ocorrer caso a transição para um determinado estado seguinte ocorra. No caso do agente de *MM*, a função de recompensa é o próprio retorno r_{t+1} do agente.

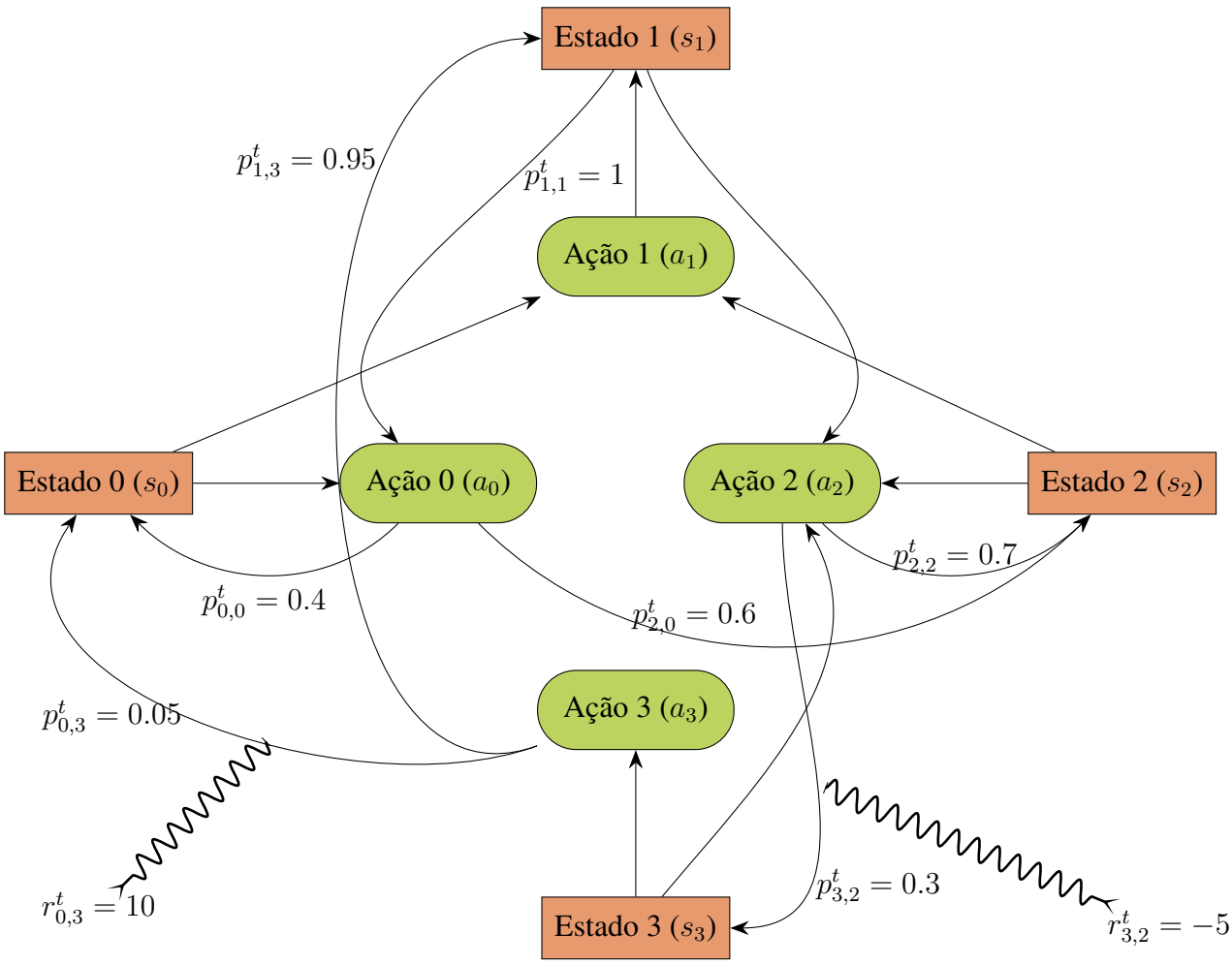

Figure 3: Processo de Decisão de Markov com 4 estados e ações (MDP)

Utilizando a definição do agente como um Processo de Decisão de Markov, podemos simular o ambiente e encontrar uma política de decisão de preços ótimas. As técnicas de Aprendizado por Reforço serão usadas para essa tarefa, e abaixo o problema inicial de otimização do agente será traduzido à notação usada na área de *RL* que parte das *MDPs* de modo a permitir o uso de algoritmos computacionais de otimização para *market making*.

- Trajetória (τ): é a sequência de estados observados e ações tomadas ao longo do tempo. No contexto de *MM*, uma trajetória consiste em uma série de estados do mercado $s_t = (o_t, L_t)$ seguidas da ação em cima desse estado $a_t = \{(\delta_t(p_i), Q_i) \mid \forall i\}$. Essas trajetórias representam a jornada do agente no mercado financeiro, incluindo suas ações e interações com o ambiente.

$$\tau = (s_0, a_0, s_1, a_1, \ldots, s_T, a_T)$$

  Onde s_t é o estado no tempo t, e a_t é a ação tomada no tempo t.

- Política (π): função que mapeia o estado atual (*spreads* δ e quantidades Q) para a escolha de ações (ofertas de compra e venda). Através de algoritmos de otimização de decisão (*Policy Optimization* e *Q-Learning*), nosso objetivo é encontrar uma política ótima que permita ao agente tomar decisões que maximizem seus retornos no mercado. Essa política é fundamental para determinar como o agente se comporta em diferentes situações de mercado.

$$\pi(s) \rightarrow a$$

  Essa função determina como o agente toma decisões em diferentes estados.

- Função de Valor (V): estima o valor esperado acumulado que o agente pode obter ao seguir a política π a partir de um estado inicial. No contexto do agente de *market making*, V depende do preço de venda p e da quantidade q executada, bem como da política do agente. Através do Aprendizado por Reforço, podemos calcular V para avaliar quão bom é um estado, o que orienta o agente na seleção de ações que maximizam seu desempenho global.

$$V(s_0) = \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t \cdot r(s_t, \pi(s_t))\right]$$

Onde γ é o fator de desconto que pondera as recompensas futuras e geralmente $0 < \gamma \leq 1$. A função valor considera a expectativa de retorno sobre todas as possíveis trajetórias do agente.

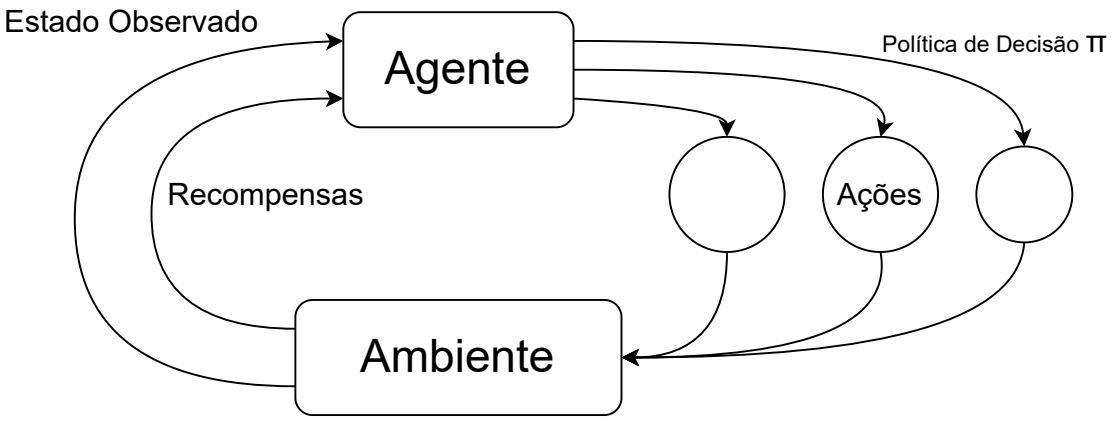

Figure 4: Autômato do Agente sob o paradigma de Aprendizado por Reforço

O agente passa a ser representado pela nova tupla $(\tau, \pi, \mathcal{V})$ a partir de um processo estocástico atrelado $\mathbf{MDP}$. O paradigma de Aprendizado por Reforço permite que o agente se adapte a mudanças nas condições de mercado ao longo do tempo, e que a política escolhida considere o impacto das transações do próprio agente sobre o mercado, tornando-o mais resiliente a flutuações causadas pelas próprias ações. Uma vez treinado, o agente pode ser usado para tomar decisões em tempo real no mercado financeiro, proporcionando uma vantagem competitiva para instituições financeiras e o avanço da pesquisa em finanças quantitativas e aprendizado de máquina, abrindo novas possibilidades para o desenvolvimento de estratégias de negociação mais eficientes e robustas. A contribuição da pesquisa proposta será tornar o agente adverso ao risco noturno, inserindo-se também uma restrição adicional, de que ao final do dia não haja exposição a riscos de mercado. Existem algumas alternativas para formalizar matematicamente essa restrição:

1. No final do dia, o agente não pode ter nenhum ativo em posição:

$$\sum_{b=1}^{B_t} q_b - \sum_{a=1}^{A_t} q_a = 0 \tag{6}$$

2. No final do dia, se houver alguma posição restante, o agente precisa *headgear*[3] sua exposição ao risco ao participar em outros mercados abertos no momento, abordagem que chamamos de *market making* **simultâneo**.

[3]De maneira simplificada, o *hedge* consiste em comprar ou vender ativos que tenham uma exposição ao risco oposta aos riscos da carteira atual, de modo a equilibrar a posição.

3.3 *Market making* simultâneo

Ao considerarmos a situação em que o MM aplica a sua estratégia em diversos mercados simultaneamente observamos um aumento da complexidade, mas também das alternativas para lidar com riscos envolvidos - chamamos essa situação de MM **simultâneo** ou **multivariado**.

Para tal, a função valor V_T do agente passa a ser a soma de todas posições parciais $V_{T,k}$, onde k é uma bolsa onde o agente possui ações:

$$V_t \quad = \quad \sum_{k=1}^{N} V_{T,k}$$
$$o$$

O objetivo principal (5) continua o mesmo, e mantém-se a restrição (6). Contudo, surgem novas alternativas para proteção da carteira durante a noite:

1. O agente pode avaliar o risco global da carteira, e incluir um único ativo de proteção contra o risco global ao final do dia.

2. Se o ativo for negociado em múltiplas bolsas (chamado também de ativo *co-listed*), o agente pode continuar a negociação deste em outra bolsa caso uma delas esteja fechada.

Considerando um cenário em que haja ações em *co-listing*, surge a possibilidade de criação de estratégias mais sofisticadas, permitindo a implementação dos itens mencionados acima.

4 Trabalhos prévios

A pesquisa sobre a minimização de risco overnight usando Aprendizado por Reforço (RL) para estratégias de market making está inserida em um contexto mais amplo de estudos que exploram a aplicação do RL em finanças quantitativas. Nesta seção, levantamos uma lista inicial de trabalhos relacionados que ajudaram a moldar e fundamentar a proposta deste projeto.

A pesquisa formal do problema do *Market Making* (MM) foi iniciada pelo estudo de Marco Avellaneda e Sasha Stoikov, em seu trabalho seminal de 2008 (Avellaneda and Stoikov, 2008). Eles abordaram o problema do MM sob certas suposições relacionadas aos processos de chegada de ordens de compra e venda. No entanto, é importante observar que, nesse cenário, eles não impuseram a restrição de que o inventário do *market maker* no final do dia fosse diferente de zero, ou seja, $q_T \neq 0$.

Como um dos resultados, Avellaneda e Stoikov conseguiram definir uma estratégia ótima para o MM, que se baseia em um cálculo cuidadoso das cotações de compra e venda em resposta às chegadas de ordens de mercado. Esta estratégia foi derivada dentro de um quadro teórico e matemático bem definido, oferecendo uma proposta clara sobre como um *market maker* pode otimizar seu desempenho em um livro de ordens.

Ao longo do tempo, diversos autores começaram a relaxar algumas das hipóteses feitas por Avellaneda e Stoikov, tornando o cenário de MM mais genérico e realista. Esse avanço na literatura expandiu as possibilidades de modelagem e análise de estratégias de MM em ambientes mais complexos e dinâmicos. Alguns exemplos são:

- "Optimal Market Making with Limited Risk" (Guéant, 2017): Este estudo aborda especificamente o problema do market making sob a perspectiva da minimização do risco. Os autores desenvolvem um modelo de market making que leva em consideração restrições de risco e investigam como otimizar a estratégia de market making enquanto limitam o risco associado.

- "High-frequency trading in a limit order book" (Avellaneda and Stoikov, 2008): Este estudo investiga as estratégias de trading de alta frequência em um livro de ordens de limite. Embora não aborde diretamente o uso do RL, fornece insights valiosos sobre o funcionamento de mercados eletrônicos e os desafios enfrentados pelos market makers, incluindo a gestão de risco e a necessidade de ajustar os preços rapidamente.

Nos últimos anos, o uso do paradigma de Aprendizado por Reforço se tornou mostrou extremamente útil para tarefas mais complexas, de jogos à medicina (Kaelbling et al., 1996). Aplicações específicas do método de *RL* no estudo de problemas de *MM* receberam grande atenção por alguns autores recentemente, notavelmente em:

- "Reinforcement Learning Approaches to Optimal Market Making" (Gašperov et al., 2021): Este estudo fornece uma visão abrangente das aplicações do Aprendizado por Reforço em market making. Os autores demonstram como o RL pode ser usado para ajustar dinamicamente os preços de compra e venda em resposta às condições do mercado. Eles destacam a eficácia do RL em otimizar o retorno ajustado ao risco em comparação com estratégias tradicionais.

- "Reinforcement Learning for Market Making in a Multi-agent Dealer Market" (Ganesh et al., 2019): Este artigo oferece uma visão detalhada de como o RL pode ser aplicado em um ambi-

ente de mercado com vários agentes, semelhante ao cenário do mundo real. Os autores demonstram que um agente de RL pode aprender a adaptar suas estratégias de market making em resposta às ações de outros agentes e às condições do mercado, incluindo a gestão de risco.

Ao revisar esses trabalhos relacionados, podemos observar que, apesar dos tratamentos teóricos bem elaborados, a situação real do uso do MM não se encaixa nas limitações impostas pelas pesquisa:

- nos mercados financeiros reais, os agentes operam de uma forma mais complexa que assumido nas pesquisas: quase todos usam estratégias MM simultâneas;
- as alternativas de proteção e as restrições de posicionamento, especialmente numa estratégia simultânea, são bem mais abrangentes que na literatura atual.

Em situações reais, diferente da proposta de Avellaneda and Stoikov (2008), não é possível encontrar uma estratégia ótima de forma analítica. O uso de técnicas de Aprendizado por Reforço em estratégias de market making oferece um potencial significativo para melhorar a eficiência das operações financeiras e mitigar os riscos associados, especialmente o risco overnight.

5 Objetivos

A pesquisa tem como objetivo central a criação de uma estratégia de market making para mercados financeiros de alta frequência. Nesse contexto, planejamos obter tanto um método de otimização para a política de preços de compra e venda, como também a criação de um ambiente de simulação em tempo real do livro de ofertas limite e de outros agentes.

A pesquisa focará em aplicações de técnicas de aprendizado de reforço para modelar o comportamento do agente de market making e obter a política ótima. Isso inclui a definição de cadeias de Markov de estado, de recompensa e de decisão. Em cenários onde a calibragem de parâmetros é necessária, serão realizados ajustes dos mesmos ao longo do tempo, de modo a levar em consideração as condições do mercado e as expectativas do agente.

Como conclusão da pesquisa, pretendemos realizar uma avaliação abrangente do desempenho da estratégia de market making, incluindo mas não limitado a análise do retorno da carteira ao longo do tempo, levando em consideração custos de transação e flutuações nos preços dos ativos. Avaliaremos também a eficácia da estratégia sob a restrição de risco *overnight* máximo e o impacto das ordens

geradas para zerar a carteira na liquidez do mercado. Após o treinamento do agente, realizaremos também uma análise comparativa entre a estratégia de market making desenvolvida com técnicas de aprendizado por reforço e estratégias tradicionais de mercado, especificamente de *price-taking*. Isso nos permitirá destacar as vantagens e desvantagens da abordagem de aprendizado de reforço e comparar os resultados obtidos quantitativamente.

6 Metodologia

O andamento da pesquisa buscará seguir a seguinte metodologia, especificamente o plano de atividades descrito — dentro dos prazos estipulados — visando concluir o objetivo comentado acima. Uma série de métricas serão utilizadas para estabelecer a eficácia do agente na tarefa proposta, assim como também comparações à outras estratégias de *price-taking*.

6.1 Planejamento de Atividades

A execução do projeto será dividida nas seguintes etapas:

1. Pesquisa Bibliográfica

   Realizar uma revisão abrangente da literatura relacionada a Aprendizado por Reforço (RL), market making e estratégias de minimização de risco em finanças quantitativas além das referências iniciais, usando bancos de dados de periódicos, como o CAFe e Arxiv. Realizar uma análise estatística dos principais algoritmos, políticas de aprendizado e bancos de dados utilizados em pesquisas anteriores. Compreender as tendências atuais e os desafios enfrentados na aplicação do *RL* em estratégias de *trading*.

2. Coleta e Processamento de Dados

   Coletar dados históricos do mercado financeiro relevantes para o estudo tendo como referência as bases de dados divulgadas na literatura. Adaptar os dados para o formato do paradigma de Aprendizado por Reforço.

3. Definição do Ambiente de Simulação

   A partir dos dados obtidos e bibliotecas de simulação, preparar um ambiente de que represente com precisão as condições do mercado de alta frequência, incluindo a dinâmica do motor do

livro de ordens, a volatilidade e o processo de chegada de ofertas. Analisar o framework de simulação mais adequado à tarefa (*OpenAI Gym*, *RLlib*, *garage*, etc.).

4. Escolha do algoritmo de Aprendizado por Reforço

   Selecionar algoritmos de *RL* adequados para a tarefa de market making e restrições adicionais (especificamente a minimização do risco), com base na revisão da literatura e na taxonomia dos algoritmos existentes (baseado em modelo, livre de modelo, entre outros).

5. Treinamento do Agente de *RL*

   Iniciar o treinamento do agente de *RL* no ambiente simulado. Definir as funções de recompensa e as métricas de desempenho relevantes para a minimização do risco overnight. Monitorar o progresso do treinamento e ajustar os hiperparâmetros conforme necessário.

6. Avaliação do Desempenho

   Avaliar o desempenho do agente de *RL* em cenários de simulação, e realizar análises estatísticas em cima dos hiperparâmetros usados. Comparar o desempenho do agente com estratégias de mercado tradicionais. Realizar ajustes no agente de *RL* com base nos resultados da avaliação de desempenho. Otimizar a estratégia de market making para alcançar um equilíbrio entre lucro e minimização de risco.

7. Divulgação e Publicação

   - Com a avaliação de desempenho finalizada, produzir um artigo científico expondo o agente criado assim como as métricas de avaliação obtidas. Preparar o artigo para submissão em conferências ou workshop na área de finanças quantitativas e aprendizado de máquina.

   - Disponibilizar os códigos do algoritmo, dados e ambiente de simulação em meios de exposição de código aberto (GitHub ou GitHub Pages). Os resultados e o código desenvolvido serão expostos usando a ferramenta de versionamento *Git*, assim como instruções para uso e detalhes de implementação do agente no formato *Markdown*.

6.2 Cronograma

Com base nas tarefas enumeradas na Seção 6.1, é mostrado na Tabela 1 o cronograma a ser executado durante a realização deste projeto.

Table 1: Cronograma das atividades.

Fases	Meses											
	1	2	3	4	5	6	7	8	9	10	11	12
Pesquisa Bibliográfica	x	x	x	x								
Coleta de Dados			x	x								
Ambiente de Simulação			x	x	x	x						
Definição do Agente					x	x	x	x				
Treinamento do Agente							x	x	x	x		
Métricas de Desempenho								x	x	x		
Produção do Artigo								x	x	x	x	x

6.3 Resultados Esperados e Métricas de Avaliação

A análise dos resultados da estratégia de *market making* será realizada diretamente em cima dos retornos obtidos pelo agente, diários e acumulados, assim como a volatilidade do portfólio e impacto de mercado das operações para zerar a carteira até o fechamento do pregão. A avaliação das decisões tomadas pelo agente ao longo do tempo e seus impactos na carteira de ativos incluem mas não se limitam à:

Desempenho: Avaliar o valor da carteira ao longo do tempo, considerando todas as transações, incluindo custos de transação e flutuações nos preços dos ativos. Comparar as métricas de desempenho mencionadas da estratégia MM com estratégias que não inserem ofertas limite no livro de ordens (chamadas de estratégias de *price-taking*), destacando desempenho melhora ou não com relação ao agente desenvolvido.

Minimização do Risco Overnight: Ao mesmo tempo, a estratégia deve minimizar o risco overnight, mantendo exposições limitadas a movimentos adversos de preços após o fechamento do mercado. A estratégia deve contribuir positivamente para a liquidez do mercado, facilitando a execução de negócios para outros participantes buscando lucro. Isso, no entanto, deve ser equilibrado com considerações de custos de transação, incluindo comissões de corretagem e outros custos associados à execução de negócios.

Explicabilidade e Interpretabilidade do Agente: Analisar as decisões do agente MM em relação à inserção e ajuste das ofertas de compra e venda, considerando a adaptação às expectativas de mercado. O agente deve levar em consideração as distribuições de quantidades executadas dado determinados preços. Serão utilizados testes estatísticos para verificar se o grafo de transições obtido pelo agente efetivamente reflete os processos do mercado real.

References

Avellaneda, M. and Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3):217–224.

Bakshaev, A. (2020). Market-making with reinforcement-learning (sac). *Quantitative Finance*.

Ganesh, S., Vadori, N., Xu, M., Zheng, H., Reddy, P., and Veloso, M. (2019). Reinforcement learning for market making in a multi-agent dealer market.

Gašperov, B., Begušić, S., Šimović, P. P., and Kostanjčar, Z. (2021). Reinforcement learning approaches to optimal market making. *Mathematics*, 9(21):2689.

Guéant, O., Lehalle, C.-A., and Fernandez-Tapia, J. (2012). Dealing with the inventory risk: a solution to the market making problem. *Mathematics and Financial Economics*, 7(4):477–507.

Guéant, O. (2017). Optimal market making.

Kaelbling, L. P., Littman, M. L., and Moore, A. W. (1996). Reinforcement learning: A survey.

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

Selser, M., Kreiner, J., and Maurette, M. (2021). Optimal market making by reinforcement learning. *Quantitative Finance*.