UNIVERSIDADE FEDERAL DE OURO PRETO (UFOP)

INSTITUTO DE CIÊNCIAS SOCIAIS APLICADAS (ICSA)

DEPARTAMENTO DE CIÊNCIAS ECONÔMICAS (DEECO)

BACHARELADO EM CIÊNCIAS ECONÔMICAS

XAIDER GOMES BRITO

OLIGOPÓLIO E O ALGORITMO Q-LEARNING

MARIANA / MG

2025

XAIDER GOMES BRITO

OLIGOPÓLIO E O ALGORITMO Q-LEARNING

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Ciências Econômicas no Instituto de Ciências Sociais Aplicadas (ICSA) da Universidade Federal de Ouro Preto (UFOP) - *Campus* Mariana, como requisito parcial para obtenção do Título de Bacharel em Ciências Econômicas.

Orientador: Prof. Dr. Martin Harry Vargas Barrenechea.

MARIANA / MG

2025

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

B862o Brito, Xaider Gomes.
Oligopólio e o algoritmo Q-Learning. [manuscrito] / Xaider Gomes Brito. - 2025.
64 f.

Orientador: Prof. Dr. Martin Harry Vargas Barrenechea.
Monografia (Bacharelado). Universidade Federal de Ouro Preto. Instituto de Ciências Sociais Aplicadas. Graduação em Ciências Econômicas .

1. Algorítmos computacionais. 2. Aprendizado do computador. 3. Cournot, A. A. (Antoine Augustin), 1801-1877. 4. Economia. 5. Linguagem de programação (Computadores). 6. NetLogo (Linguagem de programação de computador). 7. Oligopólios. I. Barrenechea, Martin Harry Vargas. II. Universidade Federal de Ouro Preto. III. Título.

CDU 004.5

Bibliotecário(a) Responsável: Essevalter De Sousa - Bibliotecário Coordenador
CBICSA/SISBIN/UFOP-CRB6a1407

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
INSTITUTO DE CIENCIAS SOCIAIS E APLICADAS
DEPARTAMENTO DE CIENCIAS ECONOMICAS

FOLHA DE APROVAÇÃO

Xaider Gomes Brito

Oligopolio e o algoritmo Q-learning

Monografia apresentada ao Curso de Ciências Econômicas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciências Econômicas

Aprovada em 7 de abril de 2025

Membros da banca

Prof. Dr.- Martin Harry Vargas Barrenechea - Orientador- Universidade Federal de Ouro Preto (UFOP)

Prof. Dr.- Carlos Eduardo da Gama Torres - Universidade Federal de Ouro Preto (UFOP)

Prof. Dr.- Luccas Assis Attílio - Universidade Federal de Ouro Preto (UFOP)

Martin Harry Vargas Barrenechea, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 07/04/2025

Documento assinado eletronicamente por **Martin Harry Vargas Barrenechea**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 07/04/2025, às 18:19, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **0891486** e o código CRC **8CDA9C3F**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: (31)3557-3835 - www.ufop.br

AGRADECIMENTOS

A jornada até a conclusão deste trabalho não foi solitária, e sou imensamente grato a todos que, de alguma forma, contribuíram para que esse momento se tornasse possível.

Primeiramente, agradeço aos meus pais, cujo apoio incondicional e incentivo foram fundamentais em cada etapa dessa caminhada.

Um agradecimento especial ao meu irmão, Tayler, por estar sempre ao meu lado, acreditando em mim e me dando forças nos momentos mais desafiadores. Sua presença foi essencial para que eu seguisse em frente.

Também expresso minha profunda gratidão ao meu professor orientador, Martin, cuja paciência, conhecimento e dedicação foram indispensáveis para a realização deste trabalho. Seu apoio e direcionamento fizeram toda a diferença na construção deste projeto.

A todos que, direta ou indiretamente, contribuíram para que eu chegasse até aqui, meu sincero obrigado.

Resumo

Este trabalho investiga os comportamentos estratégicos emergentes em mercados oligopolistas simulados por meio de modelagem baseada em agentes (MBA) e algoritmos de aprendizado por reforço, com foco no Q-learning. Utilizando a plataforma NetLogo, foram simuladas firmas autônomas que competem em um ambiente com demanda linear e custos marginais nulos, tomando decisões adaptativas baseadas em recompensas. O estudo busca compreender se, ao longo do tempo, os agentes tendem a convergir para o equilíbrio de Cournot, formar colusões tácitas ou adotar comportamentos instáveis. As simulações foram conduzidas com variações paramétricas sistemáticas e analisadas estatisticamente no ambiente R. Os resultados mostram que a configuração do ambiente e os parâmetros de aprendizagem — como número de agentes e taxa de exploração — influenciam fortemente os padrões emergentes, oscilando entre regimes cooperativos e não cooperativos. A pesquisa contribui para o debate sobre racionalidade limitada em economia industrial, mostrando como algoritmos de aprendizado podem replicar dinâmicas reais de mercado.

Palavras chave: Oligopólio. Q-learning. Modelagem Baseada em Agentes. Economia Computacional. Aprendizado por Reforço. Cournot. Simulação Multiagente. NetLogo.

ABSTRACT

This work investigates the emergence of strategic behavior in simulated oligopolistic markets through agent-based modeling (ABM) and reinforcement learning algorithms, focusing on Q-learning. Using the NetLogo platform, autonomous firms were simulated competing in an environment with linear demand and zero marginal costs, adapting their production strategies over time based on endogenous rewards. The study aims to understand whether agents tend to converge toward Cournot equilibrium, develop tacit collusion, or exhibit unstable and non-strategic behaviors. Simulations were carried out with systematic parametric variations and the results were analyzed statistically using R. The findings show that both the environment configuration and the learning parameters—such as the number of agents and the exploration rate—strongly influence the emergent patterns, which fluctuate between cooperative and non-cooperative regimes. This research contributes to the debate on bounded rationality in industrial economics by showing how learning algorithms can replicate real-world market dynamics.

Keywords: Oligopoly. Q-learning. Agent-Based Modeling. Computational Economics. Reinforcement Learning. Cournot. Multiagent Simulation. NetLogo..

Sumário

1 INTRODUÇÃO

Nas últimas décadas, o avanço das tecnologias computacionais e dos métodos baseados em simulação transformou profundamente o modo como fenômenos econômicos complexos podem ser investigados. Em especial, a *Economia Computacional Baseada em Agentes* (ACE — *Agent-Based Computational Economics*) emergiu como uma abordagem promissora para modelar sistemas econômicos descentralizados, nos quais múltiplos agentes interagem de forma adaptativa, muitas vezes sem alcançar equilíbrios analíticos convencionais. Conforme destacam TESFATSION, 2006 e EPSTEIN, 1999, a capacidade de simular comportamentos emergentes a partir de microfundamentos programáveis não apenas amplia a fronteira metodológica da economia, como também inaugura uma nova epistemologia científica baseada em experimentação virtual.

A relevância desta pesquisa reside precisamente na aplicação dessas ferramentas computacionais para o estudo de mercados oligopolistas, estruturas econômicas caracterizadas por interdependência estratégica entre poucas empresas e dinâmicas concorrenciais frequentemente distantes dos modelos clássicos de concorrência perfeita. Trabalhos como os de AXELROD, 1997 e ARTHUR, 1994 demonstram que agentes com racionalidade limitada podem, por meio de aprendizado iterativo, gerar padrões de comportamento que desafiam os pressupostos da teoria dos jogos tradicional. Nesse contexto, o uso de algoritmos de *Reinforcement Learning* (RL), como o Q-learning WATKINS; DAYAN, 1992, permite investigar como firmas simuladas ajustam suas estratégias ao longo do tempo em ambientes estocásticos e não cooperativos.

Além de sua relevância metodológica, o presente trabalho também possui importância prática e teórica ao buscar compreender se agentes autônomos, ao interagirem repetidamente em ambientes oligopolistas, tendem à competição à la Cournot, à colusão tácita ou a comportamentos intermediários. A literatura recente aponta evidências ambíguas nesse sentido XU, 2021, SHI; ZHANG, 2020, sugerindo que o design do algoritmo de aprendizado e os parâmetros do ambiente podem influenciar decisivamente os resultados. Diante disso, esta pesquisa propõe um arcabouço experimental capaz de isolar variáveis estruturais e comportamentais, fornecendo evidências replicáveis sobre os padrões emergentes em mercados simulados.

O tema deste trabalho é a análise de comportamentos estratégicos emergentes em mercados oligopolistas simulados, utilizando modelagem baseada em agentes e algo-

ritmos de aprendizado por reforço. O estudo se concentra em um ambiente construído na plataforma NetLogo, no qual firmas heterogêneas competem por meio de decisões de produção, adaptando suas estratégias ao longo do tempo com base em recompensas endógenas. O escopo da pesquisa foi delimitado ao modelo de Cournot com demanda linear e custos marginais nulos, a fim de proporcionar clareza analítica e comparabilidade com benchmarks teóricos.

A partir desse recorte, a pergunta que orienta esta pesquisa é: *Agentes econômicos simulados, ao aprenderem por reforço em um ambiente oligopolista descentralizado, tendem a convergir para o equilíbrio de Cournot, para padrões de colusão tácita ou para estratégias não estratégicas e instáveis?*

O objetivo geral da pesquisa é investigar, por meio de simulações computacionais, como firmas simuladas baseadas em Q-learning ajustam suas decisões de produção em ambientes oligopolistas e quais padrões estratégicos emergem dessas interações.

Para alcançar esse objetivo, os seguintes objetivos específicos foram estabelecidos:

1. Modelar um ambiente oligopolista simulado com base na estrutura do modelo de Cournot, incorporando múltiplas firmas autônomas com racionalidade limitada;
2. Implementar o algoritmo Q-learning na lógica decisória dos agentes, parametrizando aspectos como taxa de exploração, fator de desconto e tempo de aprendizagem;
3. Executar experimentos computacionais variando o número de agentes, o grau de exploração e a intensidade da demanda, observando os efeitos dessas variáveis sobre os lucros, preços e padrões de produção;
4. Comparar os resultados simulados com os referenciais teóricos do equilíbrio de Nash-Cournot e da colusão perfeita, identificando convergências, desvios e padrões intermediários.

Com base na literatura teórica e nos estudos prévios, esta pesquisa trabalha com duas hipóteses principais: (i) em ambientes com poucos agentes e baixa taxa de exploração, é possível a emergência de comportamentos colusivos implícitos, mesmo sem comunicação entre firmas; (ii) quanto maior a taxa de exploração e o número de firmas, maior a divergência em relação ao equilíbrio de Nash e maior a instabilidade nas decisões dos agentes.

Metodologicamente, trata-se de uma pesquisa experimental-computacional de natureza quantitativa, baseada em simulações conduzidas por modelagem baseada em agentes. O experimento foi implementado na plataforma NetLogo, utilizando a extensão `qlearningextension` para operacionalizar o algoritmo Q-learning. As simulações

foram executadas com variações paramétricas sistemáticas por meio do módulo *BehaviourSpace*, e os dados resultantes foram analisados estatisticamente no ambiente R com o suporte do pacote `tidyverse`. O processo seguiu um desenho laboratorial virtual, com controle rigoroso sobre os parâmetros do ambiente e análise posterior dos comportamentos emergentes.

A literatura sobre aprendizado em jogos oligopolistas ainda apresenta posições divergentes sobre os efeitos da racionalidade limitada. Modelos clássicos como os de Cournot pressupõem equilíbrio estratégico e expectativas consistentes. No entanto, autores como AXELROD, 1997, EPSTEIN, 1999 e LEBARON, 2006 apontam que, sob racionalidade limitada, os agentes podem desenvolver padrões cooperativos sem necessidade de imposição institucional. Outros, como MYATT; WALLACE, 2015 e LIAN; ZHENG, 2021, destacam o papel da instabilidade e da heterogeneidade de expectativas, sugerindo que o aprendizado descentralizado nem sempre leva à convergência.

Por outro lado, estudos empíricos com simulações — como os de WALTMAN; KAYMAK, 2008 e XU, 2021 — mostram que algoritmos como o Q-learning podem replicar resultados próximos à colusão, dependendo das configurações de exploração. Ainda há debate sobre até que ponto essas convergências são robustas ou artefatos do modelo. Ao abordar esses pontos de vista distintos, este trabalho busca contribuir com evidências adicionais, promovendo uma síntese crítica entre as abordagens comportamentais e computacionais da teoria dos jogos.

O referencial teórico do trabalho está dividido em três partes principais. A primeira discute os fundamentos da economia computacional, com ênfase na abordagem baseada em agentes e na epistemologia das simulações. A segunda parte apresenta os conceitos centrais do aprendizado por reforço, incluindo a formulação matemática dos Processos de Decisão de Markov e o funcionamento do algoritmo Q-learning. A terceira parte explora o modelo de Cournot como arcabouço teórico para a modelagem dos mercados simulados.

Os resultados obtidos indicam que os agentes simulados, ao aprenderem por reforço, podem adotar padrões de produção compatíveis tanto com o equilíbrio de Cournot quanto com regimes de colusão tácita, dependendo da configuração do ambiente e dos parâmetros de aprendizado. Em cenários com maior número de firmas e menor taxa de exploração, observou-se tendência à coordenação estratégica emergente, mesmo sem comunicação entre os agentes. Já em ambientes altamente exploratórios, os comportamentos tornaram-se mais voláteis, com lucros inferiores aos benchmarks teóricos.

A discussão dos resultados será conduzida com base na comparação entre os valores simulados e os referenciais teóricos esperados, incluindo equilíbrio de Nash, colusão perfeita e estratégias não cooperativas.

2 Referencial Téorico

2.1 Economia Computacional: Evolução e Fundamentos Teóricos

A **economia computacional** surgiu como uma resposta à crescente complexidade dos sistemas econômicos modernos e à limitação dos modelos analíticos tradicionais em capturar dinâmicas adaptativas, heterogeneidade de agentes e comportamentos fora do equilíbrio. Trata-se de um campo interdisciplinar que combina economia, ciência da computação, teoria dos sistemas complexos e inteligência artificial, com o objetivo de compreender e simular fenômenos econômicos por meio de algoritmos e experimentações virtuais.

Segundo TESFATSION, 2006, a economia computacional pode ser definida como o uso de métodos computacionais para representar explicitamente os processos microeconômicos que dão origem a padrões macroeconômicos observáveis. Diferentemente da economia analítica clássica, que parte de pressupostos agregados e agentes representativos, a economia computacional permite simular diretamente a interação entre múltiplos agentes heterogêneos, com racionalidade limitada e capacidade adaptativa.

Uma das vertentes mais relevantes da economia computacional é a chamada **Economia Computacional Baseada em Agentes (ACE — *Agent-based Computational Economics*)**, cujo principal expoente é a Modelagem Baseada em Agentes (MBA ou ABM). Essa abordagem permite representar mercados como sistemas descentralizados compostos por agentes autônomos, cada qual tomando decisões com base em regras simples e aprendendo com o ambiente.

A ACE tem raízes nas ideias de Herbert Simon (racionalidade limitada), Friedrich Hayek (ordem espontânea) e na teoria da complexidade, conforme desenvolvida no Santa Fe Institute a partir da década de 1990. Autores como ARTHUR, 1994 e AXELROD, 1997 foram fundamentais para consolidar a abordagem computacional como alternativa viável à modelagem de equilíbrios estáticos.

Ainda nos anos 1970, Thomas Schelling SCHELLING, 1971 já havia demonstrado como padrões coletivos de segregação poderiam emergir a partir de regras individuais simples, antecipando muitas ideias posteriormente formalizadas pela MBA. Esses trabalhos pioneiros abriram caminho para a simulação de comportamentos econômicos complexos como formação de preços, coordenação estratégica, aprendizado e falhas de mercado

com base na interação iterativa de agentes programáveis.

Nos anos 2000, a disseminação de ferramentas como o NetLogo, Repast e MASON tornou a MBA acessível a uma nova geração de pesquisadores, ampliando seu uso em economia industrial, finanças computacionais e teoria dos jogos evolutivos. Trabalhos como os de LEBARON, 2000 na área de finanças e EPSTEIN, 1999 na modelagem social reforçaram o valor explicativo da abordagem.

No contexto da economia industrial, a MBA permite representar empresas com racionalidade limitada, que ajustam suas estratégias com base em experiências passadas. A incorporação de algoritmos de *machine learning*, como o aprendizado por reforço, potencializa esse tipo de simulação, permitindo estudar mercados dinâmicos com adaptação estratégica ao longo do tempo. Como observam BRENNER, 2006, esse tipo de modelagem é especialmente útil para testar hipóteses comportamentais e identificar equilíbrios emergentes em contextos descentralizados.

Assim, a presente pesquisa se insere em uma tradição consolidada da economia computacional, ao aplicar a modelagem baseada em agentes e algoritmos de aprendizado por reforço para investigar a dinâmica de mercados oligopolistas simulados. Ao utilizar o NetLogo como plataforma e o Q-learning como mecanismo de aprendizado, o trabalho contribui para o avanço metodológico e analítico da ACE, explorando como estratégias racionais limitadas podem convergir (ou não) para equilíbrios econômicos conhecidos.

Embora a economia computacional já se destaque por seu potencial analítico, sua força metodológica reside justamente na forma como a simulação computacional se consolida como uma nova epistemologia científica. A seguir, discutimos essa perspectiva metodológica em maior profundidade, destacando o papel das simulações como instrumentos gerativos de descoberta, especialmente no estudo de sistemas econômicos complexos.

2.2 Simulação Computacional como Epistemologia: Entre Teoria e Experimento

Ao longo da evolução da ciência econômica, predominou uma abordagem dual entre teoria analítica e métodos empíricos. No entanto, a crescente complexidade dos sistemas sociais contemporâneos — marcados por não linearidade, interação estratégica e adaptação dinâmica — desafia os limites dessas duas frentes tradicionais. Nesse contexto, a simulação computacional emergiu como uma **terceira via metodológica**, capaz de gerar conhecimento por meio de experimentação virtual controlada.

Segundo AXELROD, 1997, a simulação computacional pode ser entendida como uma forma *gerativa* de ciência. Em vez de apenas descrever fenômenos ou testá-los empiricamente, esse método busca *produzi-los*, a partir da implementação de regras comportamentais mínimas nos agentes simulados. Esse processo permite observar padrões

emergentes e mecanismos causais que seriam inalcançáveis por métodos tradicionais.

Essa visão também é reforçada por EPSTEIN, 1999, que propõe a abordagem da **ciência gerativa**, na qual a explicação de um fenômeno requer a capacidade de "crescer"o fenômeno em questão dentro de um modelo computacional. Ao reproduzir comportamentos macroeconômicos a partir de micro-regras programadas, torna-se possível testar hipóteses causais, observar equilíbrios dinâmicos e explorar cenários contra-factuais de forma segura e controlada.

No campo da economia, essa abordagem abre espaço para o desenvolvimento de modelos que incorporam **racionalidade limitada**, **heterogeneidade de agentes**, **interações estratégicas locais** e **aprendizado adaptativo**. Em vez de assumir agentes representativos e equilíbrios a priori, os modelos baseados em agentes permitem que essas propriedades surjam do comportamento descentralizado e iterativo entre os participantes do sistema.

Além disso, o caráter **experimental da simulação computacional** é evidente: parâmetros podem ser controlados, condições iniciais variáveis podem ser testadas, e intervenções virtuais podem ser executadas para avaliar suas consequências, tudo isso com reprodutibilidade e transparência. Conforme observam GILBERT; TROITZSCH, 2005, esse paradigma permite construir "laboratórios virtuais"para estudar sistemas sociais e econômicos que seriam de difícil manipulação no mundo real.

Portanto, a presente pesquisa se insere nesse contexto metodológico, adotando a simulação computacional não apenas como ferramenta de apoio, mas como estrutura epistemológica. Ao combinar Modelagem Baseada em Agentes com algoritmos de aprendizado por reforço, o estudo visa investigar, de forma gerativa e exploratória, como estratégias racionais limitadas emergem e evoluem em mercados oligopolistas simulados.

2.3 Sistemas Complexos e Modelagem Baseada em Agentes (MBA)

2.3.1 Modelagem Baseada em Agentes

Neste capítulo, exploramos como ferramentas computacionais podem ser aplicadas ao estudo de sistemas complexos, com ênfase na Modelagem Baseada em Agentes (MBA). Essa abordagem é particularmente promissora para representar a dinâmica de sistemas compostos por múltiplas entidades autônomas, como é o caso de mercados com empresas heterogêneas. A seguir, serão discutidos o conceito central da MBA, suas características principais, as ferramentas que viabilizam sua implementação, suas aplicações em diferentes áreas e, por fim, sua relevância para esta pesquisa.

A Modelagem Baseada em Agentes (MBA ou ABM, do inglês *Agent-Based Modeling*) é uma abordagem computacional voltada à simulação de sistemas formados por

múltiplos agentes autônomos. Cada agente é capaz de tomar decisões, interagir com outros agentes e adaptar-se ao ambiente em que está inserido. Essa estrutura permite representar diretamente a heterogeneidade e a descentralização das ações, características fundamentais dos sistemas complexos.

Os agentes, enquanto unidades básicas da MBA, são definidos como entidades com comportamentos específicos, programadas para responder a regras predefinidas, estímulos ambientais e interações sociais. Segundo WILENSKY; RAND, 2015, essa configuração modular fornece uma base flexível para representar sistemas sociais e naturais. A simulação dessas interações gera dados observáveis, tanto quantitativos quanto visuais, o que contribui para uma análise rica e detalhada dos fenômenos modelados.

Essa capacidade de representar interações locais entre agentes torna a MBA particularmente eficaz em contextos onde estruturas macroeconômicas emergem da microinteração entre os componentes. Por meio da implementação de regras simples no nível individual, é possível observar padrões coletivos complexos que dificilmente seriam previstos por modelos tradicionais.

Até aqui, discutimos as bases conceituais da MBA e seu potencial para representar sistemas descentralizados. A seguir, abordamos as ferramentas computacionais que viabilizam sua aplicação prática.

A disseminação da MBA está diretamente associada ao avanço das ferramentas computacionais. Softwares como o NetLogo destacam-se por sua interface acessível e pela facilidade com que permitem construir modelos interativos. Além disso, a integração com algoritmos de aprendizado por reforço, como o Q-learning, amplia o realismo dos modelos ao permitir que os agentes aprendam com base nas recompensas obtidas ao longo do tempo.

Esse tipo de integração é especialmente valioso em contextos econômicos, onde agentes tomam decisões recorrentes com base em experiências anteriores. BRENNER, 2006 destaca que modelos de aprendizado por reforço são adequados para representar esse tipo de comportamento adaptativo, pois dispensam conhecimento explícito do ambiente e favorecem a exploração de estratégias.

A abordagem baseada em agentes encontra aplicação em diversas áreas, como economia, ecologia, ciências sociais, mobilidade urbana e dinâmicas eleitorais. BONABEAU, 2002 enfatiza que sua força reside na capacidade de capturar comportamentos emergentes originados a partir de decisões descentralizadas. Agentes que interagem localmente, com base em regras simples, podem produzir padrões coletivos complexos, o que reforça o potencial explicativo da MBA.

Além disso, EPSTEIN, 1999 argumenta que a ABM não deve ser vista apenas como uma ferramenta descritiva, mas como uma abordagem *gerativa*, capaz de produzir

explicações causais por meio da simulação de comportamentos individuais. Essa perspectiva é fundamental para investigações que buscam compreender os mecanismos internos de sistemas complexos.

Em suma, vimos até aqui como a MBA oferece uma estrutura poderosa para simular agentes autônomos, suas interações e os efeitos emergentes. Na sequência, discutiremos os limites e os cuidados necessários para aplicação dessa abordagem.

Apesar de seu potencial, a MBA não representa uma solução universal. A adequação da abordagem depende das especificidades do sistema modelado e dos objetivos analíticos do pesquisador. BRENNER, 2006 afirma que não existe um modelo *per se* capaz de representar todos os tipos de aprendizado, ressaltando a necessidade de ajustar a modelagem aos contextos específicos.

DUFFY, 2006 complementa essa visão, destacando a importância de calibrar os modelos com base em evidências empíricas e no comportamento efetivo de agentes reais. Na economia experimental e comportamental, por exemplo, a MBA permite testar como regras simples de aprendizado podem gerar resultados compatíveis com os dados observados.

Encerradas as discussões conceituais e metodológicas, passamos agora à análise da relevância da MBA para a presente pesquisa.

Neste trabalho, a escolha pela Modelagem Baseada em Agentes justifica-se por sua capacidade de representar, com fidelidade e dinamismo, as interações entre agentes econômicos em ambientes complexos. A MBA permite simular empresas heterogêneas com racionalidade limitada, interações locais e efeitos emergentes, características presentes no contexto de concorrência monopolista no comércio internacional.

Ao implementar agentes com aprendizado por reforço, torna-se possível analisar como decisões empresariais evoluem ao longo do tempo em resposta à interação com outras firmas e com o ambiente competitivo. Isso permite observar padrões de comportamento, testar hipóteses e explorar cenários alternativos de mercado.

Portanto, ao incorporar aspectos como heterogeneidade, autonomia e interação entre agentes, a modelagem baseada em agentes oferece uma base robusta para a compreensão dos fenômenos econômicos investigados neste estudo.

2.3.2 NetLogo como Ferramenta para Modelagem Baseada em Agentes

O *NetLogo* é uma linguagem de programação e um ambiente de desenvolvimento dedicado à construção e execução de modelos baseados em agentes. Desenvolvido por Uri Wilensky no *Center for Connected Learning and Computer-Based Modeling*, da Northwestern University, o NetLogo foi concebido para facilitar a simulação de sistemas complexos

compostos por múltiplas entidades interativas WILENSKY; RAND, 2015.

Uma de suas principais virtudes é a interface acessível, que permite aos usuários criar, visualizar e manipular modelos de maneira interativa. Essa característica favorece a experimentação com diferentes cenários, a análise de sensibilidade e o acompanhamento do comportamento do sistema ao longo do tempo, tornando o NetLogo uma plataforma ao mesmo tempo didática e robusta para análise computacional PRISMA – Centro de Física Teórica e Computacional, 2010.

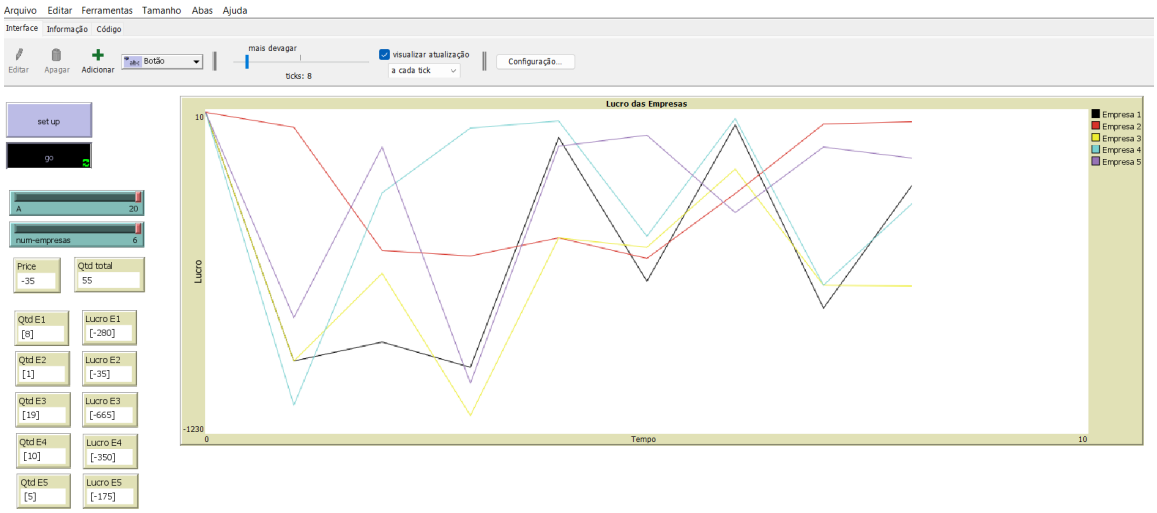
A Modelagem Baseada em Agentes (MBA) encontra no NetLogo um ambiente especialmente apropriado, pois ambas as abordagens compartilham o foco na representação descentralizada e dinâmica de sistemas complexos. A capacidade da plataforma de representar agentes autônomos — com regras próprias de decisão e interação com o ambiente — está em consonância com os fundamentos teóricos da MBA, o que contribui para sua popularização como ferramenta de modelagem RAILSBACK; GRIMM, 2020; BANOS et al., 2015.

De acordo com WILENSKY; RAND, 2015, o NetLogo foi projetado para viabilizar a simulação de fenômenos sociais, naturais e artificiais programáveis, incorporando elementos como heterogeneidade, adaptação e comportamento emergente. Neste contexto, discute-se nesta seção o papel do NetLogo na aplicação da MBA, com foco no presente estudo.

RAILSBACK; GRIMM, 2020 descrevem o NetLogo como uma plataforma especializada em MBA, destacando sua linguagem de alto nível, abordagem conceitual acessível e ferramentas integradas para experimentação automatizada. Essa combinação o torna particularmente eficaz para a construção de modelos com múltiplos agentes interativos, mesmo em contextos de grande complexidade. GILBERT; TROITZSCH, 2005 reforçam essa visão ao apontar que simulações computacionais, especialmente as baseadas em agentes, permitem investigar fenômenos sociais e econômicos que não poderiam ser analisados por métodos tradicionais.

Para ilustrar as funcionalidades descritas, a Figura **??** apresenta a interface do NetLogo durante a execução do modelo utilizado neste estudo, destacando os principais elementos de controle e visualização.

Figura 1 – Interface do NetLogo durante a execução do modelo de simulação econômica. No canto superior esquerdo estão os botões de inicialização e execução do modelo. À esquerda, encontram-se os controles de parâmetros (como número de empresas e coeficiente A), além dos monitores de variáveis-chave como preço, quantidade total e lucro individual. À direita, um gráfico dinâmico exibe o lucro das empresas ao longo do tempo, permitindo observar o desempenho comparativo de cada agente.



Fonte: Elaborado pelo autor.

No escopo desta pesquisa, os agentes são representados exclusivamente por empresas, cada uma dotada de autonomia para escolher suas ações com base em sua própria experiência acumulada. Essas empresas funcionam como entidades independentes, com estados internos específicos e estratégias de decisão baseadas em aprendizado por reforço. Essa estrutura permite simular processos adaptativos em ambientes competitivos, nos quais as decisões evoluem a partir da interação entre os agentes e das recompensas obtidas ao longo do tempo.

Diferentemente de outros modelos desenvolvidos no NetLogo que fazem uso de representações espaciais explícitas como grades de células ou mapas bidimensionais, o presente estudo não atribui relevância analítica à localização dos agentes. O foco está na dinâmica de aprendizado e nas estratégias adotadas ao longo de interações repetidas. Assim, o NetLogo é empregado como plataforma de simulação e controle das interações econômicas entre agentes racionais limitados, sem necessidade de visualização espacial.

A autonomia dos agentes refere-se à sua capacidade de tomar decisões de forma independente, guiando-se por seus próprios objetivos e estratégias. No caso das empresas simuladas neste trabalho, essa autonomia se traduz na busca pela maximização do lucro, com base em ajustes estratégicos decorrentes das recompensas recebidas. Como destacam RAILSBACK; GRIMM, 2020, essa característica favorece o surgimento de comportamentos adaptativos, nos quais os agentes reavaliam e modificam suas ações em resposta a mudanças no ambiente ou nas ações dos demais agentes.

Apesar de não utilizar recursos gráficos complexos, o NetLogo permite acompanhar em tempo real a evolução das decisões dos agentes e a dinâmica de seu aprendizado, o que facilita a análise de comportamentos emergentes. Essa funcionalidade contribui para uma compreensão mais profunda dos processos de ajuste estratégico e para a identificação de padrões coletivos que possam surgir ao longo das simulações.

Mesmo sem visualizações espaciais, o acesso contínuo a variáveis-chave como produção, lucro e recompensas acumuladas torna os resultados mais acessíveis e interpretáveis. Essa interação com os dados gerados facilita a detecção de tendências, instabilidades ou padrões adaptativos, enriquecendo a análise dos processos de aprendizado multiagente.

Outro aspecto relevante é a capacidade do NetLogo de realizar análises de sensibilidade de forma eficiente. Modelos baseados em agentes geralmente dependem de múltiplos parâmetros como taxas de aprendizado, fatores de desconto, níveis de exploração e número de agentes que influenciam significativamente os resultados. A possibilidade de alterar esses parâmetros de forma controlada permite investigar como o sistema responde a diferentes configurações BANOS et al., 2015.

Esse tipo de análise é fundamental para avaliar a robustez do modelo e compreender a influência relativa de cada variável sobre os padrões observados. Ao ajustar sistematicamente os parâmetros e observar suas consequências nos resultados, é possível identificar os fatores mais impactantes, contribuindo tanto para o refinamento do modelo quanto para o entendimento das dinâmicas simuladas.

Além disso, o NetLogo constitui um ambiente seguro para experimentações virtuais. No contexto desta pesquisa, a simulação de estratégias de produção e aprendizagem em ambientes oligopolistas permitiu explorar, de forma controlada, os efeitos de diferentes configurações de mercado — como variações no número de empresas ou nos níveis de exploração adotados pelos agentes. Essa abordagem fornece subsídios para a análise de possíveis desdobramentos de comportamentos estratégicos e suas implicações de longo prazo.

Ao permitir a comparação entre cenários distintos e a testagem de hipóteses sob diferentes condições, o NetLogo contribui para uma tomada de decisão mais informada, mesmo que em nível experimental. Esse recurso é particularmente útil em áreas como economia industrial e teoria dos jogos evolutivos GILBERT; TROITZSCH, 2005.

Entretanto, é importante reconhecer os desafios envolvidos na construção e análise de modelos baseados em agentes. A calibração e validação dos parâmetros podem ser processos sensíveis, exigindo múltiplas iterações até que os resultados se mostrem consistentes. Além disso, os comportamentos emergentes nem sempre são intuitivos, podendo demandar técnicas estatísticas ou visualizações complementares para sua adequada interpretação.

Conforme proposto por WILENSKY; RAND, 2015, o processo de modelagem no NetLogo é iterativo e incremental, integrando conceitos teóricos à experimentação computacional. Esse processo pode ser resumido nas seguintes etapas, adaptadas ao contexto desta pesquisa:

1. **Definição do problema**: Identificação do fenômeno a ser modelado. Neste estudo, trata-se da dinâmica competitiva entre empresas em um mercado oligopolista, com foco no aprendizado estratégico ao longo do tempo via algoritmos de reforço.

2. **Construção do modelo conceitual**: Estruturação de uma representação conceitual dos agentes, suas ações possíveis, estados observáveis e regras de interação. As empresas ajustam suas quantidades de produção com base em recompensas obtidas.

3. **Implementação computacional**: Codificação do modelo no NetLogo, com uso da extensão `qlearningextension`, que viabiliza a simulação de agentes baseados em aprendizado por reforço.

4. **Experimentação e validação**: Execução de simulações sob diversas configurações paramétricas, com vistas à verificação da robustez dos resultados e à comparação com expectativas teóricas.

5. **Análise dos resultados**: Avaliação dos dados gerados, com destaque para o desempenho dos agentes sob diferentes níveis de exploração (ϵ). A comparação entre cenários permitiu investigar a estabilidade, convergência e sensibilidade dos comportamentos simulados.

O NetLogo destaca-se, portanto, como uma ferramenta acessível e poderosa para a construção e análise de modelos baseados em agentes. Sua linguagem intuitiva, aliada à interface interativa e ao suporte conceitual integrado, torna-o adequado tanto para iniciantes quanto para pesquisadores experientes.

Ao permitir a simulação de comportamentos emergentes e a exploração de fenômenos complexos, o NetLogo facilita a análise de sistemas dinâmicos. A possibilidade de simular em tempo real e ajustar parâmetros dinamicamente favorece a identificação de padrões, a compreensão das interações entre agentes autônomos e a formulação de hipóteses fundamentadas.

Em um cenário científico que demanda cada vez mais abordagens computacionais para lidar com a complexidade dos sistemas sociais e econômicos, o NetLogo consolida-se como uma plataforma versátil, contribuindo significativamente para a experimentação, a análise e a tomada de decisão baseada em evidências.

2.3.3 Análise Estatística dos Dados com RStudio e BehaviourSpace

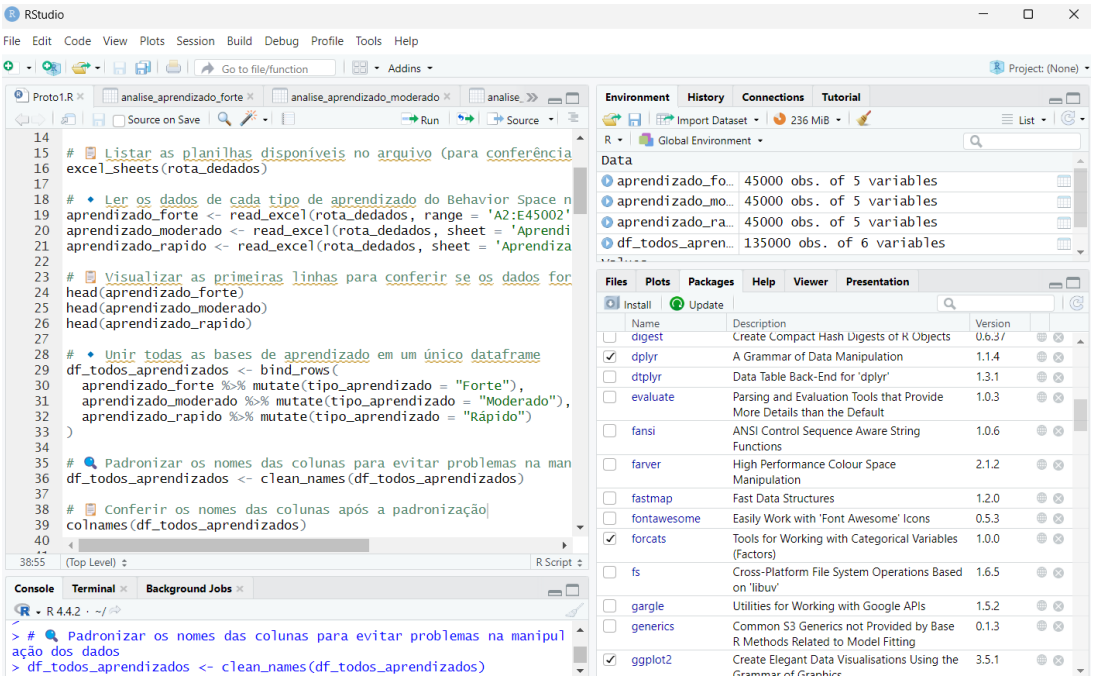
A execução de modelos baseados em agentes, especialmente em plataformas como o NetLogo, resulta na geração de um grande volume de dados experimentais. Esses dados são essenciais para a avaliação da robustez dos modelos, da consistência dos comportamentos simulados e da relação entre parâmetros e variáveis de interesse. No entanto, sua análise requer ferramentas capazes de organizar, tratar e sintetizar essas informações de maneira estatística e visual. Nesse sentido, a utilização de recursos como o BehaviourSpace e da linguagem R, em conjunto com o ambiente RStudio, mostrou-se fundamental para a etapa analítica desta pesquisa.

O BehaviourSpace é uma funcionalidade nativa do NetLogo projetada para realizar experimentos computacionais com rigor e reprodutibilidade. Por meio dele, é possível definir conjuntos de parâmetros, rodar múltiplas simulações automatizadas e exportar as variáveis de interesse para posterior análise. Como destacado por WILENSKY; RAND, 2015, esse recurso permite que o pesquisador explore sistematicamente o espaço paramétrico de um modelo, capturando dados confiáveis em grande escala. No presente estudo, o BehaviourSpace foi empregado para executar cenários de simulação envolvendo diferentes parâmetros de aprendizado, variações no número de agentes e no parâmetro estrutural A, registrando ao final de cada execução estatísticas de desempenho médio das empresas modeladas.

Para tratar os dados oriundos do BehaviourSpace, optou-se pela linguagem R em conjunto com o ambiente RStudio. Segundo HORTON; KLEINMAN et al., 2015, essa combinação oferece um ecossistema completo para análise estatística, visualização gráfica e reprodutibilidade científica. O R permite realizar desde operações básicas de limpeza e organização até análises descritivas, testes estatísticos e modelagens avançadas. O RStudio, por sua vez, proporciona um ambiente integrado para desenvolvimento, documentação e execução das análises. Essa escolha metodológica permitiu sistematizar o processo analítico da pesquisa e registrar com transparência cada etapa da análise.

A Figura 2 apresenta a interface do RStudio utilizada na condução das análises, destacando a estrutura dos scripts, a visualização dos dados e os recursos de documentação integrados.

Figura 2 – Interface do RStudio utilizada na análise dos dados.



Fonte: Elaborado pelo autor

Os dados gerados pelas simulações foram importados a partir de planilhas `.xlsx` para o ambiente R e consolidados em um único *dataframe*. Etapas preliminares incluíram a limpeza das variáveis, padronização de nomes e conversão de tipos, garantindo a consistência estrutural do conjunto de dados.

A análise exploratória foi conduzida com base em estatísticas descritivas e visualizações gráficas, permitindo examinar a distribuição dos lucros médios sob diferentes configurações experimentais, notadamente, variações no número de empresas e no parâmetro de demanda. Paralelamente, foram calculados lucros teóricos com base nos equilíbrios de Nash e colusivo, estabelecendo um referencial para comparação com os resultados produzidos pelos agentes com Q-learning.

Para investigar a influência dos parâmetros do modelo sobre os resultados econômicos, estimaram-se regressões lineares sob diversas especificações, com e sem transformação logarítmica da variável dependente. Modelos com termos de interação foram testados, e erros padrão robustos foram empregados para mitigar efeitos de heterocedasticidade.

A utilização da linguagem R, em conjunto com o ambiente RStudio, contribuiu para o rigor e a transparência do processo analítico. Conforme destacam KRONTHALER; ZÖLLNER, 2021, o R oferece um ecossistema integrado que favorece a rastreabilidade dos procedimentos e a replicação dos resultados. No contexto desta pesquisa, o uso de scripts parametrizados permitiu automatizar tarefas, documentar cada etapa e facilitar a reprodutibilidade dos experimentos.

Como extensão do escopo metodológico, prevê-se a aplicação de modelos de regressão multivariada com foco na quantificação dos efeitos do nível de exploração, número de agentes e tipo de aprendizado sobre os indicadores de desempenho observados nas simulações.

2.4 Aprendizado por Reforço (RL, TD, MDP, Q-learning)

A tomada de decisões em ambientes econômicos dinâmicos, como mercados oligopolistas, envolve incertezas, feedbacks constantes e múltiplos agentes com racionalidade limitada. Nesses contextos, as empresas enfrentam o desafio de ajustar suas estratégias de produção e competição com base em interações contínuas com o ambiente e com seus concorrentes. O aprendizado por reforço (RL — *Reinforcement Learning*) apresenta-se como uma abordagem apropriada para representar esse processo adaptativo, permitindo que agentes aprendam a tomar decisões ótimas ao longo do tempo com base na experiência acumulada. Este capítulo apresenta os conceitos fundamentais do aprendizado por reforço, discutindo sua estrutura formal, desafios e, posteriormente, o algoritmo Q-learning, amplamente utilizado em simulações multiagente.

No aprendizado por reforço, um agente aprende a associar estados a ações de forma a maximizar um sinal de retorno acumulado ao longo do tempo. Entretanto, a aplicabilidade dessa abordagem enfrenta limitações importantes, sendo uma das mais relevantes a chamada *maldição da dimensionalidade*. Esse termo, introduzido por BELLMAN, 1961, descreve o crescimento exponencial do hipervolume dos dados à medida que o número de variáveis do sistema aumenta. Esse fenômeno impacta diretamente a viabilidade computacional de algoritmos de RL, dificultando a exploração exaustiva do espaço de estados e ações. Para mitigar esses efeitos, técnicas de aproximação de funções, como redes neurais, são frequentemente utilizadas, ajustando seus pesos com base nas recompensas recebidas pelo agente após a interação com o ambiente LINS, 2020.

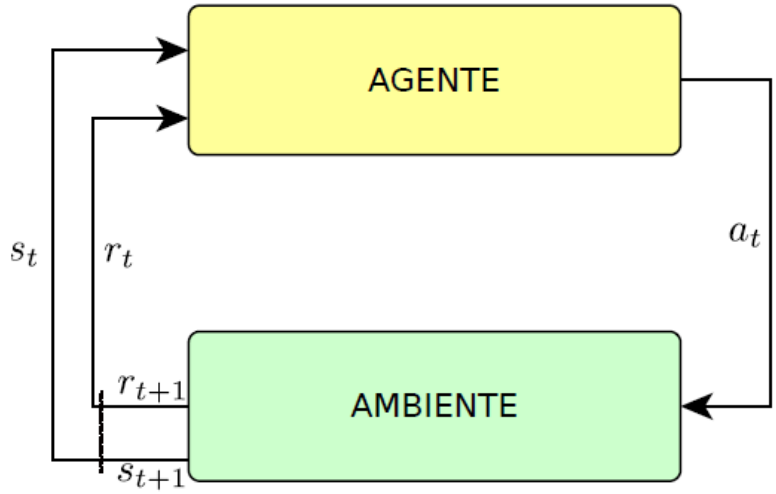
Outro desafio clássico enfrentado por algoritmos de aprendizado por reforço é o *dilema entre exploração e explotação*. Para maximizar a recompensa acumulada, o agente deve equilibrar a escolha entre explorar novas ações (com o objetivo de adquirir conhecimento sobre o ambiente) e explorar ações já conhecidas que, até então, mostraram bons resultados. Conforme discutido por BELLEMARE et al., 2016; TANG et al., 2017, o excesso de exploração pode atrasar o aprendizado, enquanto a explotação prematura pode levar a decisões subótimas e à estagnação em estratégias locais.

Adicionalmente, destaca-se o problema da *atribuição de crédito*, relacionado à dificuldade de associar recompensas futuras às ações passadas que as originaram. Como as decisões de um agente afetam não apenas o estado imediato, mas também os resultados subsequentes, torna-se desafiador identificar quais ações, em qual momento, foram res-

ponsáveis pelas recompensas recebidas. A solução desse problema é fundamental para que o agente aprenda uma política ótima π^*, que mapeie estados para ações maximizando o retorno esperado. Esse processo é formalizado como um Processo de Decisão de Markov (MDP) LINS, 2020.

A Figura 3 ilustra a dinâmica de interação entre o agente e o ambiente no contexto do aprendizado por reforço.

Figura 3 – Interação do agente com o ambiente de aprendizagem por reforço.



Fonte: SUTTON; BARTO, 2018

2.4.1 Estrutura Matemática: Processos de Decisão de Markov (MDPs)

Um **Processo de Decisão de Markov (PDM)** é um modelo matemático utilizado para descrever processos de tomada de decisão sequenciais, nos quais a dinâmica do sistema satisfaz a **propriedade de Markov**. Essa propriedade estabelece que o futuro do processo depende apenas do estado presente e da ação escolhida, sendo independente da sequência de eventos passados que levaram a esse estado BELLMAN; BELLMAN; CORPORATION, 1957; DERMAN, 1970. Formalmente, um PDM é representado por um conjunto de estados S, um conjunto de ações A, uma função de transição de probabilidades $P(s'|s,a)$, que descreve a chance de transição entre estados dada uma ação, e uma função de recompensa $R(s,a)$, que determina o retorno esperado ao executar uma ação em um determinado estado. Em cada instante de tempo t, um agente observa um estado $s_t \in S$, escolhe uma ação $a_t \in A$ seguindo uma política $\pi(a_t|s_t)$, recebe uma recompensa r_t e transita para um novo estado s_{t+1} segundo a dinâmica probabilística do ambiente. Essa formulação permite a modelagem de uma ampla gama de problemas em otimização e controle, fornecendo um arcabouço teórico para algoritmos de aprendizado por reforço.

A depender da natureza do problema, um PDM pode ser **episódico** ou **não episódico**. No caso episódico, há um estado terminal em que o processo se reinicia ao final

de cada episódio, formando sequências finitas de estados, ações e recompensas. O retorno acumulado ao longo de um episódio é definido por:

$$R_t = \sum_{k=0}^{T-1} r_{t+k+1}$$

onde T representa a duração do episódio e r_{t+k+1} é a recompensa obtida em cada etapa. Já em PDMs não episódicos, onde o processo ocorre indefinidamente, um fator de desconto $\gamma \in [0,1]$ é introduzido para ponderar a relevância das recompensas futuras, garantindo a convergência do retorno esperado:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}$$

Neste caso, se $\gamma = 0$, o agente prioriza apenas as recompensas imediatas, enquanto $\gamma = 1$ faz com que ele leve em consideração todas as recompensas futuras com igual importância BELLMAN; BELLMAN; CORPORATION, 1957. O objetivo do aprendizado por reforço é encontrar uma **política ótima** π^*, que maximize o valor esperado do retorno para todos os estados:

$$\pi^* = \arg\max_{\pi} \mathbb{E}[R \mid \pi]$$

Esse problema é resolvido utilizando técnicas de **Programação Dinâmica**, como **Iteração de Valores** e **Iteração de Políticas**, baseadas na equação de Bellman, permitindo a construção de estratégias ótimas para tomada de decisão em ambientes estocásticos PUTERMAN, 1994.

2.4.2 Funções de Valor em Aprendizado por Reforço

A capacidade de um agente aprender a tomar boas decisões depende, fundamentalmente, de sua habilidade em avaliar as consequências esperadas de suas ações. Em aprendizado por reforço, essa avaliação é formalizada por meio das chamadas **funções de valor**, que atribuem um valor numérico a cada estado (ou a cada par estado-ação), indicando o retorno esperado a partir daquela condição. Essas funções não apenas guiam o comportamento do agente durante o processo de aprendizagem, como também fundamentam os principais algoritmos de aprendizado por reforço, incluindo o Q-learning, foco deste estudo.

Em um Processo de Decisão de Markov (PDM), a função de valor-estado $V^\pi(s)$ representa a expectativa do retorno acumulado quando o agente inicia no estado s e segue uma política π. Já a função de valor estado-ação $Q^\pi(s,a)$ indica o retorno esperado ao executar a ação a no estado s, também sob a política π. Formalmente, temos:

$$V^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \mid s_t = s\right],$$
$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k+1} \mid s_t = s, a_t = a\right]$$

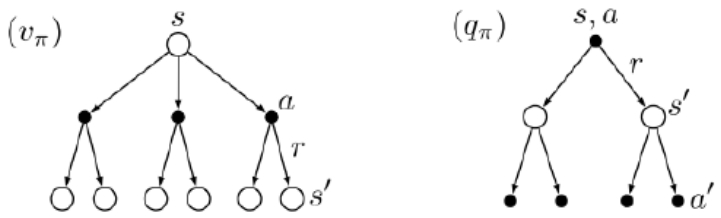
Quando o agente segue uma política ótima π^*, obtêm–se as funções $V^*(s)$ e $Q^*(s,a)$, que maximizam o retorno esperado em cada estado ou par estado–ação. Como discutido por LINS, 2020, essas funções estruturam algoritmos de decisão sequencial amplamente utilizados, inclusive em contextos econômicos, como o modelado neste estudo.

A função $Q^*(s,a)$ pode ser descrita por meio da **equação de otimalidade de Bellman**, que estabelece uma relação recursiva entre o valor de uma ação presente e os valores máximos das ações futuras:

$$Q^*(s,a) = \mathbb{E}\left[r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a') \mid s_t = s, a_t = a\right]$$

Essa equação é central para métodos baseados em iteração de valor, como o Q-learning. Ela permite que o agente atualize seus valores Q ao longo do tempo, utilizando as recompensas obtidas e as estimativas do melhor retorno possível a partir do próximo estado.

Figura 4 – Diagrama de backup para funções de valor: à esquerda, o valor do estado v_π; à direita, o valor estado-ação q_π.



Fonte: SUTTON; BARTO, 2018

Visualmente, essa propagação dos valores pode ser representada por diagramas como o da Figura 4, que ilustram como o valor presente depende das estimativas futuras mais promissoras.

Entretanto, a implementação prática dessas funções enfrenta desafios importantes. Em problemas com muitos estados e ações — como os encontrados em ambientes econômicos — manter uma tabela explícita de todos os pares (s,a) torna-se inviável. Para contornar essa limitação, são utilizados métodos de **aproximação de funções**, que estimam os valores com base em parâmetros ajustáveis, como pesos em redes neurais. Segundo BOYAN; MOORE, 1994, essa generalização é essencial para escalar o aprendizado por reforço, embora introduza novos riscos, como instabilidade e convergência imprecisa.

Além disso, diferentes estratégias podem ser utilizadas para estimar e atualizar essas funções de valor. SZEPESVáRI; LITTMAN, 1999 e LITTMAN, 2001 mostram que diversas variantes do aprendizado por reforço — incluindo versões para jogos de Markov multiagente — partem da mesma estrutura baseada em $Q(s,a)$, adaptando os métodos de atualização para diferentes cenários e objetivos.

Neste estudo, a função valor estado-ação $Q(s,a)$ será central para a modelagem do comportamento das empresas simuladas. Por meio dela, os agentes poderão aprender, de forma autônoma e descentralizada, quais estratégias de produção maximizam seus retornos em contextos de concorrência dinâmica. No entanto, a atualização eficaz desses valores exige técnicas capazes de lidar com ambientes estocásticos e aprendizado baseado em experiência. Nesse contexto, o método de Diferença Temporal (Temporal Difference – TD) emerge como uma estratégia fundamental, permitindo que os agentes ajustem suas estimativas de valor de forma incremental, à medida que interagem com o ambiente. A seguir, exploramos os fundamentos da abordagem TD e seu papel nos algoritmos de aprendizado por reforço.

2.4.3 Diferença Temporal (Temporal Difference – TD)

A aprendizagem por Diferença Temporal (TD) representa uma das abordagens mais fundamentais e eficazes do aprendizado por reforço. Seu princípio central reside na atualização incremental da função de valor com base nas diferenças entre estimativas sucessivas, dispensando o conhecimento prévio do modelo do ambiente. Essa característica torna a TD especialmente útil em contextos dinâmicos e parcialmente observáveis, nos quais o agente aprende enquanto interage com o ambiente, sem precisar conhecer explicitamente suas regras de transição.

Diferentemente dos métodos baseados em Monte Carlo — que exigem a observação de trajetórias completas para atualizar os valores —, os algoritmos TD ajustam suas estimativas após cada passo de tempo, utilizando o que é chamado de *erro TD*. Segundo SUTTON; BARTO, 2018, esse erro é definido pela diferença entre a recompensa imediata somada à estimativa do próximo estado e a estimativa atual:

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t)$$

Com base nesse erro, a função de valor é atualizada conforme a regra:

$$V(s_t) \leftarrow V(s_t) + \alpha \cdot \delta_t$$

onde α é a taxa de aprendizado, e γ o fator de desconto aplicado às recompensas futuras. Essa forma de atualização permite que o agente aprimore suas estimativas em

tempo real, ajustando sua política à medida que acumula experiência.

A aprendizagem por diferença temporal, portanto, combina os pontos fortes da programação dinâmica — ao incorporar atualizações baseadas em estimativas — com a flexibilidade dos métodos baseados em amostras, garantindo eficiência mesmo em ambientes estocásticos e sem modelo. Como resultado, ela fornece a base conceitual sobre a qual algoritmos amplamente utilizados, como o Q-learning, são construídos.

Na próxima seção, exploraremos como o Q-learning utiliza esse princípio da diferença temporal para permitir que agentes autônomos aprendam estratégias ótimas por meio da interação direta com o ambiente.

2.4.4 Q-Learning: Aprendizado por Diferença Temporal

O Q-Learning é um dos algoritmos mais conhecidos e amplamente utilizados dentro do campo do aprendizado por reforço. Sua principal característica é permitir que agentes aprendam a tomar decisões ótimas em ambientes desconhecidos por meio de tentativa e erro. Trata-se de um método baseado na ideia de Diferença Temporal (TD), no qual o agente atualiza sua função de valor a partir da recompensa imediata e da estimativa do melhor valor futuro possível.

No cerne do Q-Learning está a função valor-ação $Q(s,a)$, que representa o valor esperado de executar uma ação a em um estado s, considerando uma política de decisão ótima. Essa função é atualizada iterativamente à medida que o agente interage com o ambiente, sendo ajustada conforme as recompensas recebidas e a qualidade das decisões futuras. De acordo com WATKINS; DAYAN, 1992, esse algoritmo busca aproximar a política ótima π^*, mesmo quando o agente segue uma política exploratória. Isso torna o Q-Learning um método *off-policy*, pois a atualização dos valores não depende exclusivamente das ações tomadas durante o processo de aprendizado.

A equação de atualização do Q-Learning é dada por:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right]$$

onde α é a taxa de aprendizado, γ o fator de desconto, r a recompensa obtida, s' o novo estado alcançado, e a' a melhor ação possível nesse novo estado. Como discutido por LINS, 2020, essa formulação permite que o agente refine sua política progressivamente, ajustando sua percepção da utilidade das ações ao longo das interações com o ambiente.

Uma vantagem importante do Q-Learning está em sua flexibilidade para lidar com diferentes tipos de ambientes, desde que todos os pares estado-ação sejam suficientemente explorados. Estratégias como a política ε-gulosa contribuem para esse processo, promovendo um equilíbrio entre a exploração de novas ações e a explotação de ações

previamente avaliadas como vantajosas. Essa combinação é fundamental para a convergência do algoritmo em direção à política ótima WATKINS; DAYAN, 1992.

Para ilustrar o funcionamento do algoritmo, a Figura 5 apresenta seu pseudocódigo, reproduzido da obra de LINS, 2020. Essa representação destaca o processo iterativo de atualização dos valores $Q(s,a)$, no qual o agente ajusta suas decisões com base nas recompensas observadas ao longo do tempo SZEPESVáRI; LITTMAN, 1999.

Figura 5 – Algoritmo Q-Learning conforme apresentado por LINS, 2020.

Algoritmo 1: Algoritmo *Q-Learning*

início
  Inicialize $Q(s,a)$ aleatoriamente
  repita
    Inicialize s
    repita
      Selecione a a partir de s utilizando π derivada de Q (por ex. ε-gulosa)
      Receba a e observe os valores de r e s'
      $Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$
    até *o passo final do episódio ser atingido*;
  até *o episódio máximo ser atingido*;
fim

Fonte: LINS, 2020

Neste trabalho, o Q-Learning será utilizado como mecanismo de aprendizado para agentes autônomos que representam empresas em um ambiente econômico competitivo. Essa abordagem permite capturar como decisões de produção podem ser refinadas ao longo do tempo, com base em experiências anteriores e nas recompensas obtidas. Na seção seguinte, serão discutidas as estratégias adotadas para operacionalizar esse algoritmo no ambiente de simulação do NetLogo, considerando as especificidades da extensão `qlearningextension` empregada neste estudo.

2.5 O Modelo de Cournot como Estrutura para Simulação de Mercados

O modelo de Cournot constitui uma das estruturas teóricas mais consolidadas da microeconomia para a análise de mercados oligopolistas. Introduzido por Antoine Augustin Cournot em 1838, trata-se do primeiro modelo a formalizar matematicamente a interação estratégica entre firmas que competem por meio da escolha de quantidades produzidas, estabelecendo, assim, um marco inicial para o pensamento estratégico na teoria econômica DAUGHETY, 2006. Embora anterior ao desenvolvimento formal da teoria dos jogos, o modelo já incorporava os principais elementos dessa abordagem: racionalidade individual, interdependência de decisões e formação de expectativas. Em sua formulação

moderna, é frequentemente interpretado como um caso específico de equilíbrio de Nash em jogos simultâneos.

A estrutura matemática do modelo de Cournot pressupõe que n firmas produzem um bem homogêneo e competem em quantidade. O preço de mercado é determinado por uma função de demanda inversa linear:

$$P(Q) = a - bQ,$$

onde $Q = \sum_{i=1}^{n} q_i$ é a quantidade total ofertada. Cada firma escolhe sua produção q_i, considerando como dadas as quantidades escolhidas pelas concorrentes, com o objetivo de maximizar seu lucro π_i. Tal comportamento leva à seguinte função de reação:

$$q_i = \frac{a - c_i - bQ_{-i}}{2b},$$

em que c_i representa o custo marginal da firma i, e Q_{-i} a produção total das demais firmas. No caso simétrico de duopólio, o equilíbrio é determinado por:

$$q_1 = q_2 = \frac{a-c}{3b}, \quad Q = \frac{2(a-c)}{3b}, \quad P = \frac{a+2c}{3}.$$

Esse resultado mostra que o modelo de Cournot gera um equilíbrio intermediário entre o monopólio e a concorrência perfeita, refletindo de maneira realista a estrutura de mercado onde há poder de mercado, mas também rivalidade estratégica.

Nas últimas décadas, a literatura expandiu significativamente esse arcabouço, incorporando elementos como diferenciação de produto, informação assimétrica, expectativas heterogêneas e ajustes dinâmicos. Por exemplo, GRISáKOVá; ŠTETKA, 2022 analisam o comportamento de três firmas com expectativas distintas — ingênuas, adaptativas e racionais — e mostram que a estabilidade do equilíbrio de Cournot pode ser severamente afetada, inclusive levando a dinâmicas caóticas sob determinadas configurações. De maneira complementar, LIAN; ZHENG, 2021 propõem um modelo com ajustes contínuos nas decisões de produção, revelando que a velocidade de adaptação das firmas influencia diretamente a convergência (ou não) ao equilíbrio teórico. Por sua vez, MYATT; WALLACE, 2015 abordam a questão informacional, demonstrando como a qualidade e a origem dos sinais disponíveis — públicos ou privados — podem distorcer a alocação de recursos, gerando ineficiências tanto para os produtores quanto para os consumidores.

Esses desdobramentos teóricos reforçam a robustez do modelo de Cournot como base para experimentações computacionais, sobretudo em ambientes simulados com agentes adaptativos. Ao empregar algoritmos de aprendizado, como o Q-learning, torna-se possível investigar como firmas autônomas aprendem estratégias produtivas e interagem ao longo do tempo. A simulação computacional, nesse contexto, permite observar

se os comportamentos emergentes convergem ao equilíbrio de Nash, manifestam colusão tácita ou seguem padrões instáveis e não estratégicos.

Dada essa versatilidade analítica, o presente trabalho adota o modelo de Cournot por três razões principais. Primeiro, porque a competição ocorre via decisão de quantidade — exatamente a variável estratégica que os agentes implementam nas simulações desenvolvidas em NetLogo. Segundo, pela simplicidade estrutural do modelo e pela existência de soluções analíticas fechadas, que fornecem *benchmarks* teóricos para a comparação direta com os resultados empíricos das simulações baseadas em aprendizado por reforço. Terceiro, porque a literatura contemporânea reconhece e utiliza amplamente o modelo de Cournot como referência em estudos de economia computacional aplicada, especialmente na simulação de interações oligopolistas com agentes heterogêneos e aprendizado adaptativo.

Nesse sentido, o estudo de WALTMAN; KAYMAK, 2008 figura como um marco inicial, ao aplicar o algoritmo Q-learning em um ambiente Cournot simulado. Os autores demonstram que agentes autônomos podem, sem comunicação explícita, aprender estratégias estáveis, com resultados que oscilam entre o equilíbrio de Nash e a colusão tácita, dependendo da intensidade da exploração adotada. Posteriormente, XU, 2021 expandiu esse escopo ao testar diversas variantes do Q-learning, identificando os impactos de diferentes políticas de aprendizado sobre a estabilidade dos lucros. Entre os achados, destaca-se que políticas de exploração agressiva comprometem a convergência, ao passo que estratégias mais conservadoras favorecem resultados próximos ao ótimo coletivo.

Aspectos comportamentais também têm sido objeto de crescente interesse na literatura recente sobre mercados oligopolistas. Trabalhos como o de ALóS-FERRER, 2004 investigam a influência de fatores cognitivos e sociais, como memória e imitação, no contexto de jogos dinâmicos entre firmas competidoras. Os autores demonstram que tais elementos podem modificar significativamente a trajetória do comportamento dos agentes, levando-os a convergir para equilíbrios híbridos entre os modelos clássicos de Cournot e Walras. A profundidade da memória utilizada no processo decisório revela-se, nesse sentido, uma variável crucial, capaz de alterar as expectativas e estratégias adotadas ao longo do tempo, influenciando diretamente o tipo de equilíbrio alcançado.

Avanços mais recentes, como o estudo de SHI; ZHANG, 2020, reforçam essa perspectiva ao empregar algoritmos de aprendizado por reforço em ambientes multiagente altamente complexos. Utilizando arquiteturas baseadas em *Deep Q-Networks* (DQN), os autores analisam de forma sistemática a dinâmica de adaptação e aprendizado coletivo em mercados competitivos simulados. Suas análises revelam padrões interessantes de estabilidade e resposta estratégica entre os agentes, demonstrando que mesmo em cenários com informação incompleta ou limitada, o uso de técnicas de inteligência artificial permite capturar nuances comportamentais e trajetórias emergentes que escapariam a

modelos puramente analíticos.

Portanto, a adoção do modelo de Cournot nesta pesquisa está solidamente fundamentada tanto por seu papel histórico como estrutura canônica da teoria dos jogos aplicada à microeconomia quanto por sua relevância contemporânea no campo da economia computacional. Além de oferecer uma base teórica bem estabelecida, o modelo de Cournot destaca-se por sua compatibilidade com frameworks de simulação baseados em agentes e por sua adequação à incorporação de algoritmos de aprendizado, como o Q-learning. Essa combinação de elegância analítica, simplicidade operacional e flexibilidade computacional o torna uma escolha apropriada para investigar a emergência de padrões estratégicos em ambientes econômicos simulados.

Nas seções seguintes, será apresentada em detalhes a arquitetura computacional desenvolvida para representar esse ambiente competitivo. O foco estará na modelagem dos agentes autônomos, no papel da interação iterativa e na implementação do algoritmo Q-learning como mecanismo central de aprendizado adaptativo. Com isso, pretende-se compreender de que maneira estratégias racionais podem emergir, estabilizar-se ou divergir em função das características estruturais do ambiente e dos parâmetros de aprendizagem adotados.

3 METODOLOGIA E IMPLEMENTAÇÃO

3.1 TIPO DE ESTUDO

Este trabalho configura-se como um **estudo experimental-computacional**, de natureza **quantitativa**, fundamentado na metodologia de **simulação baseada em agentes** (*Agent-Based Modeling, ABM*) com incorporação de algoritmos de **aprendizado por reforço** (*Reinforcement Learning, RL*). A pesquisa é classificada, portanto, como um *estudo exploratório e indutivo com base empírica simulada*, cujo objetivo central é investigar a emergência de padrões estratégicos de comportamento em mercados oligopolistas sob condições de racionalidade limitada e adaptação dinâmica.

A abordagem metodológica adota o paradigma da **economia computacional baseada em agentes** (ACE — *Agent-based Computational Economics*), inserindo-se na tradição iniciada por TESFATSION, 2006 e consolidada por autores como ARTHUR, 1994, AXELROD, 1997 e EPSTEIN, 1999, onde sistemas econômicos complexos são estudados a partir da simulação de múltiplos agentes heterogêneos em interação descentralizada. Cada agente é programado para tomar decisões individuais com base em regras comportamentais mínimas, atualizadas iterativamente a partir da experiência acumulada em ambiente simulado.

O desenho da pesquisa foi orientado pelo princípio da **ciência gerativa**, conforme proposto por EPSTEIN, 1999, no qual o conhecimento sobre fenômenos socioeconômicos complexos é obtido por meio da construção de modelos computacionais capazes de "gerar"comportamentos observáveis a partir de microfundamentos programáveis. Neste caso, a simulação não serve apenas como ilustração de hipóteses teóricas, mas como *instrumento ativo de descoberta*, permitindo a replicação de fenômenos macroeconômicos a partir de comportamentos microeconômicos parametrizados.

O estudo não se baseia em dados empíricos observacionais, mas sim em **dados gerados endogenamente** pelo modelo simulado. Os agentes — representando empresas oligopolistas — atuam dentro de um ambiente competitivo virtual, no qual ajustam iterativamente suas decisões de produção em resposta aos retornos obtidos em interações anteriores. Tais decisões são mediadas por um algoritmo de aprendizado por reforço, especificamente o **Q-learning**, conforme descrito por WATKINS; DAYAN, 1992 e aplicado em estudos similares por WALTMAN; KAYMAK, 2008 e XU, 2021.

O modelo foi desenvolvido e executado na plataforma **NetLogo**, amplamente reconhecida na literatura de modelagem baseada em agentes por sua acessibilidade, poder de abstração e integração com bibliotecas de aprendizado adaptativo WILENSKY; RAND, 2015; RAILSBACK; GRIMM, 2020. A extensão `qlearningextension` foi empregada para operacionalizar a lógica algorítmica do Q-learning no interior dos agentes, mantendo a separação entre as decisões estratégicas das firmas e a infraestrutura do ambiente simulado.

A lógica subjacente ao experimento computacional pode ser descrita como um **design de simulação controlada**, no qual o pesquisador especifica previamente os parâmetros estruturais do ambiente (como número de agentes, grau de exploração, função de demanda e tempo de simulação) e executa múltiplas rodadas sob condições variáveis, a fim de observar padrões emergentes. Esses padrões são posteriormente analisados estatisticamente para identificar tendências, convergências ao equilíbrio de Nash ou possíveis manifestações de colusão tácita.

Dada sua natureza simulada, esta pesquisa também pode ser classificada como um **estudo laboratorial virtual**, em que o controle sobre as variáveis permite a testagem sistemática de hipóteses econômicas sob diferentes cenários paramétricos. O uso de simulação computacional como método científico, neste contexto, é defendido por autores como GILBERT; TROITZSCH, 2005, que destacam o papel da modelagem baseada em agentes como uma terceira via epistemológica entre dedução teórica e inferência empírica.

Portanto, o presente estudo combina rigor metodológico com flexibilidade experimental, valendo-se de ferramentas computacionais para explorar, de forma indutiva, como decisões descentralizadas de agentes com racionalidade limitada podem (ou não) conduzir a equilíbrios previsíveis em ambientes de competição estratégica. Essa configuração confere à pesquisa não apenas validade analítica, mas também um elevado grau de reprodutibilidade e transparência metodológica.

3.2 Fontes de Dados e Referenciais

Este trabalho é fundamentado em uma estrutura teórica consolidada da literatura de economia computacional, modelagem baseada em agentes e aprendizado por reforço aplicado à teoria dos jogos e mercados oligopolistas. Dado o caráter simulado da pesquisa, as **fontes de dados** utilizadas não provêm de bases empíricas reais, mas sim de **dados gerados artificialmente** por meio de simulações computacionais controladas. Esses dados são derivados diretamente do comportamento dos agentes autônomos implementados no ambiente computacional, cujas ações evoluem com base na interação dinâmica entre regras de decisão programadas e feedbacks do ambiente simulado.

Referenciais Teóricos

O referencial teórico central deste estudo está ancorado na linha de pesquisa conhecida como **Agent-Based Computational Economics** (ACE), conforme estabelecido por autores como TESFATSION, 2006, LEBARON, 2006 e FARMER; FOLEY, 2009. O núcleo da abordagem ACE reside na simulação de interações entre agentes heterogêneos, que tomam decisões com base em heurísticas ou algoritmos de aprendizado adaptativo, ao invés de supor a racionalidade plena dos modelos neoclássicos.

Para modelar a estrutura de mercado, adota-se o arcabouço teórico de **concorrência oligopolista de Cournot**, no qual firmas simultaneamente escolhem níveis de produção com o objetivo de maximizar lucros, assumindo a produção dos concorrentes como dada. O modelo é estendido para um contexto dinâmico e adaptativo, no qual as decisões das firmas não são mais estáticas, mas evoluem ao longo do tempo por meio do mecanismo de aprendizado por reforço.

Em termos de aprendizado computacional, o modelo de decisão adotado é baseado no algoritmo **Q-learning**, introduzido por WATKINS; DAYAN, 1992, que permite aos agentes aprenderem a partir da experiência, atualizando suas estratégias com base nos retornos observados das ações anteriores. Essa abordagem já foi utilizada com sucesso na literatura econômica para simular comportamentos estratégicos emergentes em jogos repetidos, como em WALTMAN; KAYMAK, 2008 e XU, 2021.

Modelos Referenciais e Bases para Implementação

A modelagem dos agentes e do ambiente competitivo segue a lógica apresentada em modelos como o de AXELROD, 1997, onde o foco recai sobre a emergência de padrões coletivos a partir de regras simples. O trabalho também dialoga com estudos empíricos e computacionais de simulações de aprendizado em estruturas de mercado, como o framework apresentado por RUST, 1997, que introduz o conceito de agentes que aprendem otimizando funções de valor em ambientes parcialmente observáveis.

Além disso, os fundamentos microeconômicos da função de demanda e estrutura de custos das firmas são inspirados nos modelos de teoria microeconômica intermediária, conforme sistematizados por VARIAN, 1992, adaptados para um ambiente computacional com função de demanda linear invertida e custos marginais constantes.

Geração e Tratamento de Dados Simulados

Todos os dados utilizados nesta pesquisa foram gerados endogenamente durante a execução dos experimentos de simulação. Em cada rodada do experimento, os agentes (firmas) escolhem níveis de produção com base em seus valores Q atualizados. As decisões

de produção resultam em combinações de quantidades totais, preços de mercado, lucros individuais e recompensas, as quais retroalimentam o algoritmo de aprendizado. Os dados gerados incluem:

- **Histórico de quantidades produzidas** por agente em cada período;
- **Preços de equilíbrio** determinados pela função de demanda agregada;
- **Lucros** obtidos por agente a cada rodada;
- **Matriz Q de aprendizado**, com os valores de retorno estimado para cada ação;
- **Estatísticas agregadas** (média, variância, desvio padrão) de variáveis econômicas ao longo do tempo.

Esses dados são armazenados automaticamente ao fim de cada simulação em estruturas de arrays no próprio ambiente NetLogo e posteriormente exportados como um arquivo CSV, dependendo do cenário experimental. A geração dos dados foi parametrizada para assegurar **reprodutibilidade** completa dos resultados, com *seeds* de aleatoriedade fixadas e documentação completa do setup experimental.

Repositórios Bibliográficos e Técnicos

Para fundamentação teórica e técnica do modelo, foram utilizados como base primária:

- O livro *An Introduction to Agent-Based Modeling* de Wilensky e Rand 2015, referência central na construção de modelos em NetLogo;
- O manual técnico da extensão `qlearningextension.nlogo`, disponibilizado pela comunidade NetLogo, para a incorporação do algoritmo Q-learning nos agentes;
- Artigos científicos revisados por pares, obtidos a partir das bases **Scopus**, **ScienceDirect**, **JSTOR**, **arXiv** e **Google Scholar**.

A integração entre os referenciais teóricos e os dados simulados permitiu construir uma base sólida para a investigação proposta, garantindo não apenas coerência conceitual, mas também **transparência metodológica e replicabilidade experimental**, condições essenciais para a validade de estudos computacionais.

3.3 Critérios de Inclusão e Exclusão de Referências

A presente pesquisa adotou critérios rigorosos para a seleção do arcabouço teórico e dos trabalhos empíricos que fundamentam o desenvolvimento do modelo proposto. A definição clara dos critérios de inclusão e exclusão visa garantir a robustez teórica, a atualidade metodológica e a relevância temática dos referenciais utilizados, alinhando-se às melhores práticas da literatura em economia computacional e modelagem baseada em agentes.

Critérios de Inclusão

Os critérios de inclusão foram estabelecidos com base em três pilares fundamentais: (i) relevância temática para a modelagem computacional em contextos econômicos; (ii) rigor metodológico e reconhecimento científico dos trabalhos; e (iii) aderência ao escopo específico da pesquisa — neste caso, a simulação de interações econômicas sob estruturas oligopolistas com agentes racionais ou adaptativos, com aplicação de aprendizado por reforço.

Foram incluídas, prioritariamente:

- Obras publicadas em periódicos com fator de impacto relevante e revisão por pares, como *Journal of Economic Dynamics and Control*, *Computational Economics*, *Nature*, *Science*, *Journal of Artificial Societies and Social Simulation (JASSS)*, entre outros.
- Trabalhos seminalmente reconhecidos na área, como os de TESFATSION e AXELROD, cujas contribuições estruturaram o campo de economia computacional e de modelagem baseada em agentes.
- Modelos de simulação em economia cuja implementação empírica e computacional tenha sido documentada de forma replicável, particularmente aqueles utilizando plataformas como NetLogo, Repast ou linguagens como Python e R.
- Pesquisas aplicadas que fizeram uso de algoritmos de aprendizado por reforço (como Q-Learning, SARSA ou Deep Q-Networks) em contextos econômicos, especialmente em jogos de Cournot e Bertrand.
- Referenciais teóricos clássicos da microeconomia (por exemplo, VARIAN), que fornecem a base analítica dos comportamentos econômicos dos agentes modelados.
- Trabalhos recentes (de 2005 em diante) que exploram a interseção entre simulação computacional, aprendizado de máquina e teoria dos jogos, com especial atenção àqueles apresentados em conferências como AAAI, NeurIPS e AAMAS.

Critérios de Exclusão

Por outro lado, os critérios de exclusão foram definidos para evitar o uso de referências que pudessem comprometer a solidez da fundamentação teórica ou desviar o foco analítico da pesquisa. Foram excluídos:

- Trabalhos de cunho opinativo, sem validação empírica, base teórica consistente ou submissão a processo de revisão científica.
- Referências que, embora relacionadas à economia computacional, abordam contextos incompatíveis com a estrutura do modelo proposto (por exemplo, simulações em mercados de trabalho ou redes sociais, quando não relacionadas a interações oligopolistas).
- Publicações anteriores ao ano de 2000, salvo quando tratam de obras seminais, de natureza teórica fundamental (como é o caso de AXELROD, WATKINS; DAYAN ou RUST).
- Trabalhos com implementações computacionais baseadas em ferramentas proprietárias não replicáveis ou com ausência de código-fonte ou documentação acessível ao público.
- Artigos que tratam de aprendizado de máquina sem conexão explícita com economia ou comportamento estratégico de agentes.

Fontes de Pesquisa e Procedimento de Triagem

As referências foram obtidas por meio de buscas estruturadas nas seguintes bases de dados e repositórios: Scopus, Web of Science, Google Scholar, RePEc, SSRN, JSTOR e arXiv. Palavras-chave como "agent-based modeling", "reinforcement learning in economics", "Cournot simulation", "Q-learning agents", "computational economics"e "NetLogo economic models"foram utilizadas em combinações booleanas.

As buscas retornaram aproximadamente 220 resultados iniciais. Após aplicação dos critérios de inclusão e exclusão, apenas 38 trabalhos foram considerados elegíveis, dos quais 27 foram efetivamente utilizados na composição do referencial teórico. A triagem foi realizada por meio da leitura de títulos, resumos e, quando necessário, do conteúdo completo, com apoio do gerenciador de referências Zotero.

Esse processo metódico garantiu a coerência entre os objetivos da pesquisa, a estrutura do modelo computacional e o estado da arte na literatura científica relevante. Ele também assegura que qualquer tentativa futura de replicação ou extensão deste estudo possa partir de um referencial bibliográfico sólido, validado e alinhado com os princípios fundamentais da ciência aberta e da economia computacional contemporânea.

3.4 Estrutura Computacional e Execução Experimental

Ambiente de Simulação

A implementação computacional do modelo foi realizada na plataforma NetLogo (versão 6.4.0), amplamente adotada em estudos de modelagem baseada em agentes pela sua expressividade e facilidade de visualização interativa. A análise estatística e o processamento dos dados gerados foram conduzidos no ambiente R (versão 4.4.2), utilizando-se pacotes como `tidyverse`, `readr` e `ggplot2`.

A execução dos experimentos seguiu um fluxo automatizado: parâmetros experimentais foram definidos via módulo BehaviourSpace, simulações foram executadas em lote, e os resultados — exportados como arquivos `.csv` — foram tratados no R para posterior análise exploratória, visualização e modelagem estatística.

Tabela 1 – Etapas do fluxo metodológico da pesquisa

Etapa	Descrição da atividade executada
1. Parametrização	Definição dos parâmetros do modelo: número de firmas, demanda máxima, taxa de exploração inicial, fator de decaimento, capacidade máxima.
2. Simulação	Execução automatizada do modelo no NetLogo com a extensão `qlearningextension.nlogo` via módulo BehaviourSpace.
3. Exportação de dados	Armazenamento dos resultados das simulações em arquivos `.csv`.
4. Análise estatística	Importação dos dados no R; tratamento e análise com os pacotes `dplyr`, `ggplot2`, `janitor`, entre outros.
5. Validação	Comparação dos resultados com os valores esperados sob equilíbrio de Cournot e colusão, além da inspeção visual e estatística dos padrões.

A princípio, o modelo foi executado com aleatoriedade não controlada. Contudo, para assegurar a reprodutibilidade futura, recomenda-se a definição explícita de sementes de aleatoriedade por meio do comando `random-seed`, o que possibilita replicar fielmente os resultados sob as mesmas condições iniciais. Esse procedimento é uma prática consolidada em experimentos computacionais e facilita a validação por pares.

O código-fonte da simulação, bem como os scripts analíticos, foram documentados e organizados em blocos funcionais, permitindo sua reutilização e extensão por outros pesquisadores. Toda a estrutura computacional foi pensada para assegurar não apenas robustez metodológica, mas também transparência e replicabilidade, valores fundamentais na pesquisa em economia computacional.

MODELAGEM E PARÂMETROS

O modelo simula um mercado oligopolista composto por firmas que ajustam suas decisões de produção ao longo do tempo com base em aprendizado por reforço. Os agentes, representados por empresas heterogêneas, interagem em um ambiente competitivo construído na plataforma NetLogo, operando segundo a lógica da Simulação Baseada em Agentes (SBA).

Cada firma escolhe discretamente entre 20 níveis possíveis de produção, definidos como ações no algoritmo de Q-Learning. A política de decisão adotada é *e-greedy*, parametrizada por uma taxa inicial de exploração (`FE`) e um fator de decaimento (`FDM`). A recompensa é dada pelo lucro, calculado a partir da diferença entre a demanda agregada ($P = A - Q$) e a produção da firma. Custos marginais foram assumidos nulos, de modo a isolar os efeitos estratégicos da competição.

O aprendizado ocorre de forma iterativa, com atualização dos valores Q ao longo dos episódios, que são reiniciados a cada 500 ciclos (`ticks`). A estrutura do modelo privilegia a simplicidade algorítmica, mantendo foco na dinâmica emergente da interação entre agentes sob restrições mínimas de racionalidade. Parâmetros como número de agentes, intercepto da demanda (`A`) e intensidade exploratória foram sistematicamente variáveis. A Tabela 2 resume os principais parâmetros testados.

Tabela 2 – Parâmetros principais utilizados nas simulações

Parâmetro	**Descrição**
`num-empresas`	Número de firmas no mercado (2, 4, 6)
`A`	Demanda máxima (10, 15, 20)
`FE`	Taxa inicial de exploração no Q-Learning (0.3, 0.5, 0.9)
`FDM`	Fator de decaimento da exploração (0.9993, 0.9995, 0.9999)
`capacidade-maxima`	Limite de produção por firma (20 unidades)
`ticks por episódio`	Duração de cada ciclo de aprendizado (500)

EXECUÇÃO DAS SIMULAÇÕES

As simulações foram conduzidas por meio do módulo BehaviourSpace do NetLogo, que permitiu a variação sistemática dos parâmetros experimentais. Para cada combinação de valores — número de firmas (`num-empresas`), intercepto da demanda (`A`), taxa de exploração (`FE`) e fator de decaimento (`FDM`) — foram realizadas 20 execuções independentes, totalizando mais de 400 simulações.

Cada execução teve duração de 88.000 ciclos temporais (`ticks`), divididos em episódios sucessivos de 500 iterações, com reinicialização do ambiente e preservação da memória de aprendizado. A estrutura foi projetada para observar padrões dinâmicos de

adaptação estratégica ao longo do tempo, incluindo trajetórias de convergência, colusão tácita e regimes oscilatórios.

Durante a simulação, variáveis-chave como produção, lucros e preço de mercado foram registradas em arquivos `.csv`, com organização por cenário experimental. Como não foi utilizado controle explícito da semente de aleatoriedade, pequenas flutuações entre execuções são esperadas. A repetição sistemática dos experimentos buscou mitigar esse efeito, assegurando robustez estatística nas análises comparativas.

Coleta e Análise de Dados

Os arquivos de saída gerados pelas simulações foram importados no ambiente R (versão 4.4.2), com apoio do RStudio como interface de desenvolvimento. A análise foi estruturada com os pacotes `readr`, `dplyr`, `tidyr`, `janitor` e `ggplot2`, permitindo um fluxo analítico replicável e modular.

O tratamento dos dados incluiu padronização de variáveis, parsing de colunas vetoriais e transformação para o formato *long*, adequado para análise estatística. Foram calculadas métricas agregadas por cenário, como lucro médio, desvio-padrão de preço e frequência de oscilações.

Os dados consolidados foram analisados por meio de modelos de regressão linear múltipla, com e sem interações, e validados com erros padrão robustos (`HC1`) para heterocedasticidade. A visualização gráfica apoiou a interpretação dos regimes comportamentais simulados, permitindo comparações com equilíbrios teóricos como Cournot ou colusão. Os scripts e dados organizados foram mantidos em repositório local, assegurando rastreabilidade e replicabilidade da análise.

Validação Metodológica

A robustez e a consistência dos resultados foram avaliadas com base em três frentes complementares de validação: verificação interna do modelo, comparação com benchmarks teóricos e análise estatística robusta dos dados simulados.

A verificação interna concentrou-se na inspeção dos outputs gerados, tanto por meio da interface visual quanto pela análise direta dos dados exportados. Foram monitoradas, ao longo dos episódios, variáveis-chave como lucros, produção agregada e preço de mercado, com especial atenção à estabilidade dos padrões e à ausência de anomalias estruturais. A reinicialização controlada do ambiente a cada 500 `ticks` e a preservação da matriz de aprendizado garantiram coerência na evolução dos agentes.

No plano teórico, os lucros médios simulados foram confrontados com valores esperados sob dois referenciais clássicos da literatura microeconômica: o equilíbrio de Nash-

Cournot e o lucro sob colusão perfeita, ambos derivados analiticamente da função de demanda linear adotada. Os resultados do modelo Q-Learning situaram-se entre os dois extremos, o que é compatível com a emergência de padrões parcialmente cooperativos sob racionalidade limitada.

Por fim, a análise estatística foi conduzida a partir de regressões lineares múltiplas, incluindo termos de interação entre parâmetros e o uso de erros padrão heterocedasticidade-consistentes (`HC1`). Essa abordagem permitiu identificar relações estatisticamente significativas entre os parâmetros de exploração, o grau de concentração de mercado e os resultados econômicos obtidos. A repetição sistemática dos experimentos para cada combinação de parâmetros assegurou poder estatístico suficiente, mesmo na ausência de controle sobre a semente de aleatoriedade.

Esse conjunto de estratégias valida não apenas a estabilidade algorítmica do modelo, mas também a plausibilidade econômica dos comportamentos emergentes observados, conferindo solidez às inferências extraídas a partir das simulações.

Considerações Finais sobre a Metodologia

A construção metodológica deste estudo combinou modelagem baseada em agentes com aprendizado por reforço, em um ambiente computacional controlado que permite explorar a dinâmica estratégica de firmas em mercados oligopolistas sob racionalidade limitada. A escolha por uma abordagem simulacional não apenas amplia o escopo de investigação teórica, como também viabiliza a observação de padrões emergentes dificilmente capturáveis por métodos analíticos tradicionais.

O modelo desenvolvido priorizou a simplicidade estrutural e a clareza comportamental dos agentes, sem abrir mão da complexidade interativa característica de sistemas econômicos descentralizados. A experimentação sistemática via variações paramétricas, associada a um pipeline de análise estatística rigoroso, assegurou a robustez dos resultados e conferiu transparência às etapas de coleta, tratamento e validação dos dados simulados.

Dessa forma, o arcabouço metodológico aqui proposto oferece mais do que uma simples estrutura operacional para a simulação de interações estratégicas entre agentes: ele constitui uma representação formalmente consistente e coerente com os objetivos analíticos desta pesquisa. Ao integrar conceitos da teoria dos jogos, técnicas de aprendizado por reforço e modelagem baseada em agentes, o modelo proporciona uma base sólida para a análise de dinâmicas emergentes em contextos oligopolistas simulados.

Essa estrutura metodológica permite capturar, de maneira sistemática, os efeitos da adaptação individual e coletiva ao longo do tempo, bem como a influência de parâmetros estruturais — como a frequência de atualização das estratégias, a intensidade da

competição e o grau de exploração versus exploração no processo de aprendizado. Assim, o modelo não apenas reproduz comportamentos teoricamente plausíveis, como também oferece um ambiente controlado e flexível para a experimentação computacional.

Consequentemente, os resultados empíricos obtidos a partir das simulações realizadas podem ser interpretados com maior confiabilidade, pois se fundamentam em uma base conceitual e técnica cuidadosamente delineada. No capítulo seguinte, serão apresentados e analisados os dados provenientes das execuções do modelo, com especial atenção às regularidades observadas, aos padrões de convergência estratégica e às implicações econômicas derivadas das trajetórias de aprendizado dos agentes.

4 Resultados e Análise Experimental

4.1 Introdução aos Resultados

Esta seção apresenta e analisa os resultados obtidos a partir das simulações computacionais conduzidas com o modelo de aprendizado por reforço baseado em Q-learning, implementado em um ambiente de mercados oligopolistas simulados por meio da plataforma NetLogo. Os experimentos foram estruturados para avaliar, sob diferentes configurações paramétricas, a capacidade dos agentes de aprender e adaptar suas estratégias de produção ao longo do tempo, em resposta aos incentivos econômicos oriundos de suas interações competitivas.

As análises visam responder diretamente à questão de pesquisa proposta — se e em que medida agentes com racionalidade limitada, guiados por mecanismos de aprendizado adaptativo, são capazes de convergir para estratégias compatíveis com equilíbrios econômicos teoricamente esperados, como o equilíbrio de Cournot ou regimes cooperativos emergentes.

Os resultados são organizados de modo a refletir os objetivos específicos delineados na etapa metodológica: i) demonstrar a robustez da implementação computacional; ii) examinar a dinâmica estratégica dos agentes sob diferentes estruturas de mercado e parâmetros de exploração; iii) avaliar padrões emergentes e trajetórias de aprendizado ao longo do tempo; e iv) comparar os resultados obtidos com benchmarks teóricos clássicos da microeconomia.

A estrutura analítica adotada combina estatística descritiva, visualização gráfica e modelagem econométrica para explorar os efeitos marginais e interativos dos parâmetros experimentais. Cada subseção a seguir discute um conjunto específico de achados, estabelecendo um diálogo sistemático entre os comportamentos observados nas simulações e os referenciais teóricos que fundamentam a pesquisa. Tal abordagem visa não apenas verificar a plausibilidade econômica dos resultados, mas também elucidar os mecanismos comportamentais subjacentes à adaptação estratégica dos agentes.

4.2 Evidência de Implementação e Execução

Esta seção apresenta evidências da correta implementação computacional do modelo de aprendizado por reforço em mercados oligopolistas simulados. O objetivo é demonstrar a funcionalidade da simulação, a coerência lógica do fluxo computacional e a automação do processo experimental via plataforma NetLogo.

Fluxo Computacional Validado

A estrutura computacional do experimento foi organizada em um pipeline modular, composto por quatro etapas principais: parametrização, execução, exportação e análise dos dados. A Tabela 1 resume as atividades realizadas em cada etapa.

Os testes de consistência do modelo foram conduzidos com base em simulações preliminares, realizadas sob diferentes combinações paramétricas. As variáveis de saída disponíveis — incluindo o número de firmas, o intercepto da função de demanda, os parâmetros do algoritmo de Q-Learning (`FE` e `FDM`) e o lucro médio das empresas ao longo dos ciclos — foram monitoradas sistematicamente para verificar a coerência interna dos resultados.

Embora a matriz de aprendizado Q e as decisões individuais de produção não tenham sido exportadas diretamente, a variável *lucro médio* ao longo do tempo forneceu uma proxy robusta do desempenho dos agentes. A análise dessa variável permitiu inferir padrões de adaptação comportamental e variações estratégicas entre os cenários, indicando que a simulação produziu trajetórias plausíveis do ponto de vista econômico.

Durante as execuções, não foram observados travamentos, falhas algorítmicas ou valores anômalos — como lucros permanentemente nulos ou irrealisticamente elevados — reforçando a estabilidade da estrutura implementada. A Figura 6 mostra uma simulação em andamento, com o gráfico do lucro das empresas oscilando ao longo do tempo, o que sugere aprendizado e ajuste de comportamento por parte dos agentes.

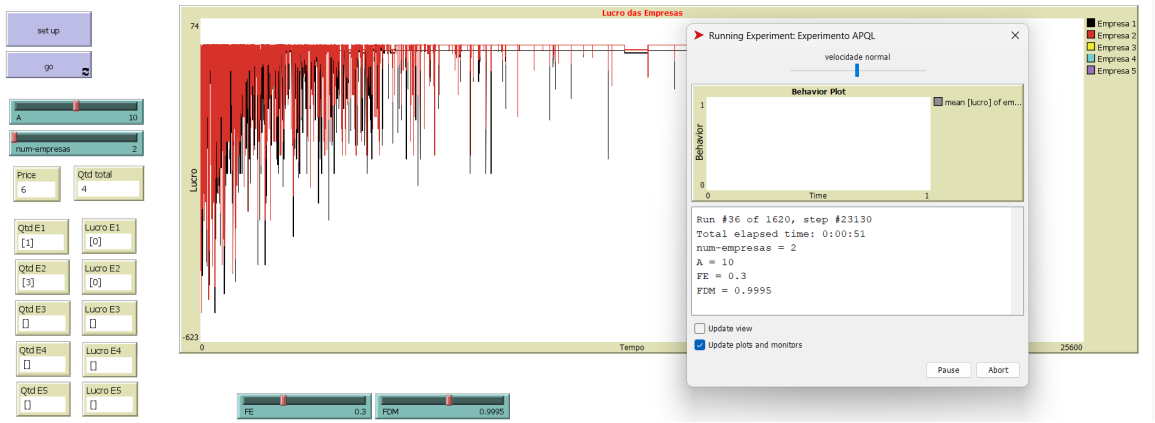


Figura 6 – Execução do experimento no NetLogo: gráfico temporal do lucro das empresas.

A automação via módulo *BehaviourSpace* funcionou conforme o esperado, permitindo a execução sistemática de múltiplos cenários e a exportação organizada dos arquivos de saída. A Figura 7 apresenta a tela de configuração do experimento, com os parâmetros variáveis e as combinações testadas.

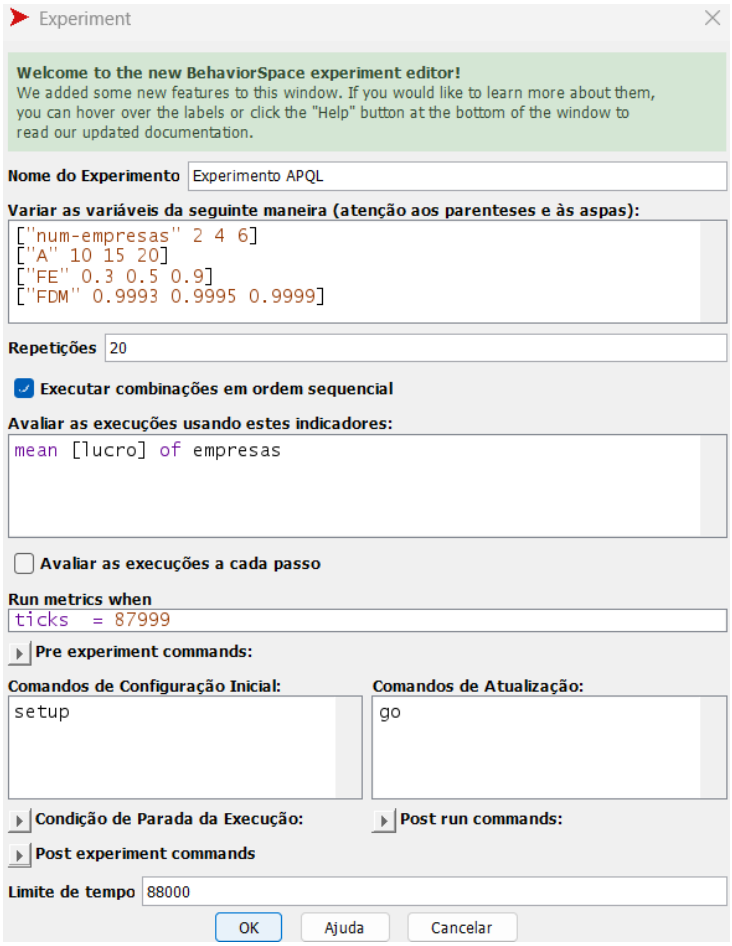


Figura 7 – Configuração do experimento no módulo BehaviourSpace.

Adicionalmente, a Figura 8 exibe o cabeçalho do arquivo `.csv` gerado pela simu-

lação, evidenciando a estrutura dos dados exportados e sua organização por parâmetros e métricas. Essa estrutura padronizada facilitou a posterior análise no ambiente R.

Column1	Column2	Column3	Column4	Column5	Column6	Column7
[run number]	num-empresas	A	FE	FDM	[step]	mean [lucro] of empresas
6	2	10	0.3	0.9993	87999	8
5	2	10	0.3	0.9993	87999	8
4	2	10	0.3	0.9993	87999	10.5
3	2	10	0.3	0.9993	87999	8
2	2	10	0.3	0.9993	87999	8
1	2	10	0.3	0.9993	87999	8
7	2	10	0.3	0.9993	87999	12
8	2	10	0.3	0.9993	87999	12.5
11	2	10	0.3	0.9993	87999	12
9	2	10	0.3	0.9993	87999	12.5
10	2	10	0.3	0.9993	87999	12
12	2	10	0.3	0.9993	87999	12.5
14	2	10	0.3	0.9993	87999	12
13	2	10	0.3	0.9993	87999	8

Figura 8 – Cabeçalho do arquivo de saída gerado pelo NetLogo.

Adicionalmente, foram conduzidas simulações de controle sob condições conhecidas, como duopólios simétricos com estratégias fixas, a fim de verificar se os resultados reproduziam padrões compatíveis com soluções teóricas clássicas, como o equilíbrio de Nash-Cournot. Essas simulações auxiliaram na validação interna do comportamento emergente do modelo.

Interface, Parâmetros e Automação

A interface gráfica do modelo, desenvolvida na plataforma NetLogo (versão 6.4.0), foi projetada para permitir controle preciso e visualização clara das variáveis experimentais. Foram incluídos sliders e switches para parametrização de variáveis críticas, como o número de firmas (`num-empresas`), o intercepto da demanda (`A`), a taxa de exploração inicial (`FE`) e o fator de decaimento da exploração (`FDM`). Monitores foram integrados à interface para acompanhar, em tempo real, variáveis endógenas como produção agregada, preço de equilíbrio e lucro médio.

A Figura 9 apresenta a interface da simulação, destacando os principais elementos de controle e monitoramento.

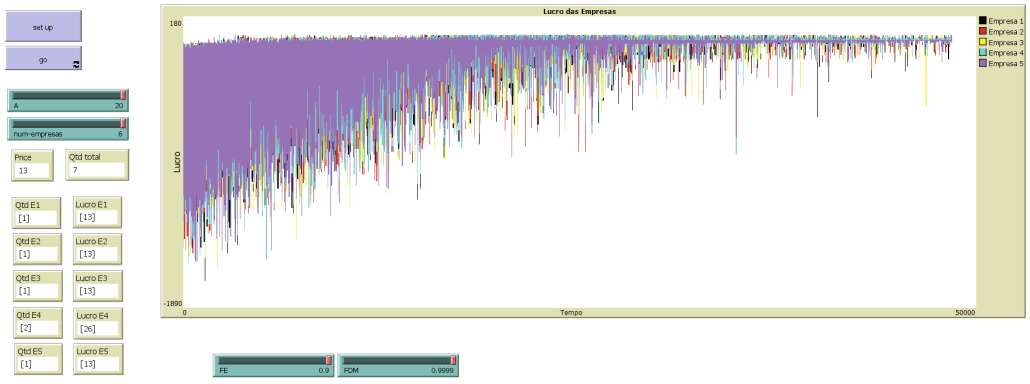


Figura 9 – Interface da simulação no NetLogo com sliders para controle paramétrico.

A execução automatizada dos experimentos foi realizada por meio do módulo *BehaviourSpace*, que possibilitou a definição de múltiplas combinações paramétricas e a execução em lote de cenários distintos. Para cada configuração experimental, foram realizadas 20 repetições com sementes pseudoaleatórias distintas, assegurando robustez estatística e controle sobre a variabilidade estocástica do modelo.

Os resultados foram exportados automaticamente em arquivos no formato `.csv`, organizados por cenário, facilitando a análise posterior. Essa automação permitiu ampla cobertura do espaço paramétrico e controle rigoroso sobre a coleta de dados. Os dados gerados foram analisados estatisticamente no ambiente R, conforme descrito na seção de metodologia analítica.

4.3 Achados Experimentais por Dimensão Investigada

4.3.1 Lucros Médios e Comparação com Referenciais Teóricos

A Tabela 3 resume os lucros médios por firma sob Q-Learning, comparando-os com os referenciais teóricos do equilíbrio de Nash-Cournot e da colusão perfeita.

Tabela 3 – Lucro por firma: Q-Learning (média ± SEM) comparado aos referenciais teóricos para diferentes níveis de demanda (A).

Dem. (A)	**Firmas**	**Q-L. (±SEM)**	**Nash (Teo.)**	**Colusão (Teo.)**
10	2	9.62 ± 0.13	11.11 (-13.5%)	12.50 (-23.1%)
	4	6.04 ± 0.02	4.00 (+50.9%)	6.25 (-3.4%)
	6	2.57 ± 0.11	2.04 (+25.8%)	4.17 (-38.4%)
15	2	24.71 ± 0.29	25.00 (-1.2%)	28.13 (-12.2%)
	4	12.27 ± 0.08	9.00 (+36.3%)	14.06 (-12.8%)
	6	8.58 ± 0.09	4.59 (+86.8%)	9.38 (-8.5%)
20	2	43.84 ± 0.59	44.44 (-1.4%)	50.00 (-12.3%)
	4	19.07 ± 0.21	16.00 (+19.2%)	20.00 (-4.6%)
	6	15.70 ± 0.06	8.16 (+92.3%)	16.67 (-5.8%)

Os dados revelam um padrão contraintuitivo: embora a teoria microeconômica preveja que o aumento no número de firmas leve a maior competição e lucros decrescentes, os agentes treinados com Q-Learning frequentemente superaram os lucros de Nash e se aproximaram dos valores colusivos — especialmente nos casos com quatro e seis firmas.

Nos cenários de duopólio, ao contrário, o Q-Learning apresentou subperformance frente aos referenciais teóricos. Isso indica que, em ambientes com baixa complexidade estratégica, os agentes podem não encontrar incentivos suficientes para explorar políticas agressivas, ficando presos a estratégias conservadoras. Essa limitação é discutida

por SUTTON; BARTO, 2018 como um desafio comum em ambientes com espaço de ação pequeno.

Já em mercados com mais competidores, a diversidade estratégica parece favorecer a emergência de coordenação implícita. Os resultados indicam padrões compatíveis com **colusão tácita emergente**, mesmo sem qualquer forma de comunicação direta entre os agentes — um fenômeno alinhado com os achados de TESFATSION, 2006 e AXELROD, 1997.

4.3.2 Estabilidade Frente à Exploração: FE e FDM

A Tabela 4 apresenta os coeficientes estimados por regressão linear múltipla com erros padrão robustos, considerando os parâmetros experimentais como variáveis explicativas dos lucros médios.

Tabela 4 – Regressão do lucro médio por firma com erro padrão robusto (HC1)

Variável	Coeficiente	Erro Padrão	t	p-valor
(Intercepto)	-522.17	578.71	-0.902	0.367
FE	0.19	0.577	0.329	0.742
FDM	525.02	578.95	0.907	0.365
Demanda (A)	2.01	0.041	49.66	< 0.001
Nº de Empresas	-4.28	0.094	-45.38	< 0.001

Nem a taxa de exploração inicial (`FE`) nem o fator de decaimento (`FDM`) apresentaram significância estatística. Isso indica que o desempenho dos agentes é relativamente estável frente à variação desses parâmetros, e sugere que a estrutura do ambiente é mais determinante para o aprendizado do que a aleatoriedade inicial. Essa observação é coerente com BELLEMARE et al., 2016 e TANG et al., 2017.

Ainda assim, os coeficientes positivos sugerem que estratégias exploratórias mais persistentes (`FDM` mais alto) podem favorecer trajetórias de aprendizado mais lucrativas — em linha com a discussão de SUTTON; BARTO, 2018 sobre o papel da exploração sustentada na superação de ótimos locais durante o aprendizado.

A Figura 10 apresenta os gráficos de diagnóstico residual do modelo estimado, incluindo as seguintes dimensões: (i) resíduos versus valores ajustados; (ii) quantis teóricos versus resíduos padronizados (Q-Q plot); (iii) escala-localização (homocedasticidade); e (iv) alavancagem versus resíduos padronizados (identificação de observações influentes).

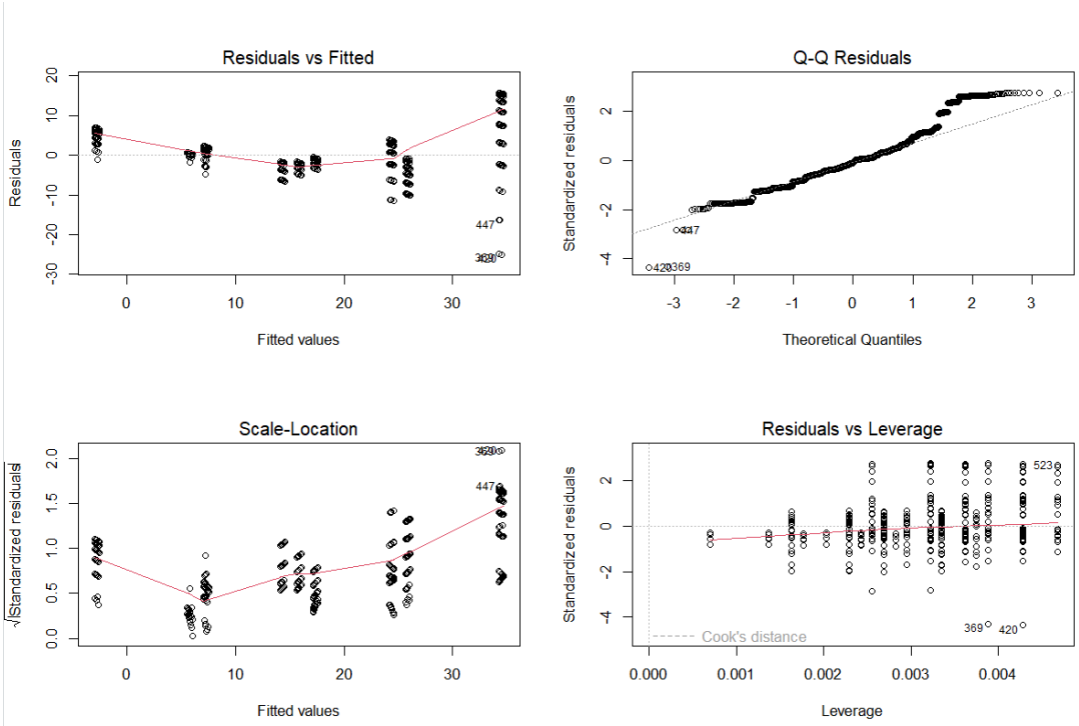


Figura 10 – Gráficos de diagnóstico do modelo de regressão linear.

A inspeção dos gráficos permite extrair algumas conclusões relevantes:

- O gráfico **"Residuals vs Fitted"** não apresenta padrões sistemáticos de curvatura ou agrupamento, indicando que o modelo não viola, de forma crítica, a suposição de linearidade entre variáveis explicativas e a variável resposta.

- O **gráfico Q-Q** indica leve assimetria à esquerda nas caudas inferiores, com alguns outliers identificados (como as observações 369, 420 e 447). Ainda assim, a maior parte dos pontos alinha-se bem à reta teórica, sugerindo uma aproximação razoável à normalidade dos resíduos.

- O gráfico de **"Scale-Location"** apresenta uma leve heterocedasticidade nas extremidades, mas sem estrutura padronizada que indique violação grave. Como precaução, erros robustos foram utilizados na regressão, mitigando o impacto estatístico de possíveis desvios.

- No **"Residuals vs Leverage"**, poucas observações possuem alavancagem elevada ou ultrapassam os limites de influência de Cook. Nenhuma delas atinge níveis que justifiquem sua exclusão ou reponderação.

Em conjunto, esses resultados reforçam a adequação do modelo aos pressupostos clássicos da regressão linear, conferindo solidez inferencial às estimativas obtidas.

4.3.3 Evidência Direta de Colusão Tácita

A Tabela 5 destaca os cenários onde o lucro médio se aproximou mais do referencial colusivo do que do equilíbrio de Nash.

Tabela 5 – Cenários com lucro médio mais próximo da colusão do que de Nash.

Firmas	**Demanda** (A)	**Lucro Q-Learning**	**Nash**	**Colusão**
4	10	6.04	4.00	6.25
4	15	12.27	9.00	14.06
6	15	8.58	4.59	9.38
6	20	15.69	8.16	16.67

Mesmo sem comunicação direta, memória ou instituição reguladora, os agentes demonstraram capacidade de desenvolver estratégias cooperativas emergentes. Trata-se de uma evidência de colusão tácita compatível com a literatura em jogos repetidos com aprendizado limitado, conforme AXELROD, 1997; TESFATSION, 2006; EPSTEIN, 1999.

4.3.4 Confronto com as Hipóteses de Pesquisa

Hipótese 1: *Ambientes com poucos agentes e baixa taxa de exploração induzem colusão tácita.*

Refutada. Os duopólios não apresentaram coordenação eficaz. Os agentes não conseguiram convergir para lucros superiores aos teóricos — evidência de que, com poucos agentes, a simplicidade estratégica pode inibir o aprendizado coletivo.

Hipótese 2: *Ambientes com alta exploração e muitos agentes são instáveis e se afastam do equilíbrio de Nash.*

Refutada parcialmente. Os ambientes com mais firmas mostraram, na verdade, maior proximidade com padrões colusivos. A taxa de exploração teve efeito estatisticamente nulo, indicando que a fragmentação pode favorecer (e não desestabilizar) a convergência cooperativa.

4.3.5 Resposta ao Problema de Pesquisa

Com base nos dados, é possível responder à pergunta central:

Agentes econômicos simulados, ao aprenderem por reforço em um ambiente oligopolista descentralizado, tendem a convergir para o equilíbrio de Cournot, para padrões de colusão tácita ou para estratégias instáveis?

Os resultados indicam que os agentes **tendem a oscilar entre o equilíbrio de Cournot e padrões de colusão tácita**, com destaque para a capacidade de aprendizado cooperativo emergente em ambientes com mais participantes. A hipótese de instabilidade não foi sustentada, e o aprendizado adaptativo se mostrou eficaz em reproduzir estruturas de mercado complexas — em linha com EPSTEIN, 1999; LEBARON, 2006.

5 CONCLUSÕES, DISCUSSÕES E PERSPECTIVAS

Quando este trabalho de pesquisa foi iniciado, partiu-se da constatação de que os modelos tradicionais da microeconomia, especialmente aqueles voltados à análise de mercados oligopolistas, enfrentam sérias limitações ao representar a complexidade dinâmica e adaptativa das interações estratégicas entre firmas reais. A justificativa teórica estava enraizada na percepção de que abordagens analíticas convencionais — como os equilíbrios de Nash em jogos estáticos — pressupõem racionalidade perfeita e conhecimento completo, o que contrasta com a realidade empírica de mercados descentralizados e sujeitos a incertezas. Diante disso, a proposta de empregar modelagem baseada em agentes (ABM) e algoritmos de aprendizado por reforço (Reinforcement Learning) se mostrou não apenas inovadora, mas metodologicamente coerente com a necessidade de explorar os padrões emergentes em sistemas econômicos complexos. O objetivo era justamente investigar até que ponto agentes autônomos, ao interagirem iterativamente e aprenderem adaptativamente, poderiam exibir comportamentos economicamente plausíveis — como colusão tácita ou convergência para equilíbrios competitivos.

O objetivo geral do estudo consistiu em investigar, por meio de simulações computacionais, como firmas autônomas baseadas em Q-Learning ajustam suas decisões de produção em ambientes oligopolistas e quais padrões estratégicos emergem dessas interações. A análise dos dados empíricos gerados nas simulações evidencia que este objetivo foi integralmente alcançado. As firmas simuladas demonstraram dinâmicas adaptativas coerentes com princípios econômicos fundamentais, oscilando entre regimes estratégicos próximos ao equilíbrio de Cournot e padrões de colusão tácita, dependendo das configurações do ambiente. A abordagem metodológica permitiu identificar com clareza as trajetórias de aprendizado e os resultados estratégicos emergentes, evidenciando que os agentes são capazes de internalizar regras de comportamento racionais, mesmo sem comunicação explícita ou coordenação institucional.

O primeiro objetivo específico era modelar um ambiente oligopolista simulado com base na estrutura do modelo de Cournot, incorporando múltiplas firmas autônomas com racionalidade limitada. Este objetivo foi plenamente alcançado por meio da construção de um ambiente em NetLogo, parametrizável e flexível, no qual as decisões de produção eram tomadas por agentes com base em recompensas econômicas endógenas. A modelagem respeitou os pressupostos do modelo Cournot clássico — demanda linear e custos

marginais nulos — permitindo assim comparabilidade com benchmarks teóricos.

O segundo objetivo era implementar o algoritmo Q-Learning na lógica decisória dos agentes, parametrizando aspectos como taxa de exploração, fator de desconto e tempo de aprendizagem. Esse objetivo também foi satisfatoriamente atendido. A extensão `qlearningextension` da plataforma NetLogo foi corretamente integrada, e os parâmetros críticos do algoritmo foram ajustáveis em tempo real. Além disso, os testes de robustez e os diagnósticos estatísticos confirmaram o funcionamento correto do algoritmo e sua influência nas decisões estratégicas dos agentes.

O terceiro objetivo consistia em executar experimentos computacionais variando o número de agentes, o grau de exploração e a intensidade da demanda, observando os efeitos dessas variáveis sobre os lucros, preços e padrões de produção. Através do módulo *BehaviourSpace*, foram realizados experimentos sistemáticos e repetíveis com diferentes configurações paramétricas. A análise dos dados, feita em ambiente R, permitiu identificar relações significativas entre essas variáveis e os padrões de comportamento emergentes, validando empiricamente a estrutura experimental desenhada.

O quarto e último objetivo era comparar os resultados simulados com os referenciais teóricos do equilíbrio de Nash-Cournot e da colusão perfeita, identificando convergências, desvios e padrões intermediários. As análises estatísticas — incluindo regressões, gráficos de diagnóstico e comparação direta com benchmarks — mostraram que os agentes Q-Learning podem, em determinados contextos, superar os lucros de Nash e se aproximar de regimes cooperativos emergentes. Esses achados sustentam a relevância do modelo experimental tanto como ferramenta analítica quanto como plataforma exploratória de hipóteses econômicas.

A pesquisa partiu da hipótese de que: (i) em ambientes com poucos agentes e baixa taxa de exploração, seria possível a emergência de comportamentos colusivos implícitos; e (ii) quanto maior a taxa de exploração e o número de firmas, maior a divergência em relação ao equilíbrio de Nash e maior a instabilidade nas decisões dos agentes. No entanto, ao longo da análise empírica, verificou-se que ambas as hipóteses foram **refutadas parcial ou completamente**. A colusão não emergiu com intensidade nos duopólios, e os ambientes com mais firmas, contrariamente à expectativa, apresentaram maior propensão à coordenação estratégica tácita. Adicionalmente, a taxa de exploração (FE/FDM) não demonstrou impacto estatisticamente significativo, indicando que a estrutura do ambiente e a heterogeneidade dos agentes são fatores mais determinantes que os parâmetros de exploração no aprendizado de estratégias.

Quanto ao problema central de pesquisa — se agentes simulados, ao aprenderem por reforço em ambientes oligopolistas, convergem para Cournot, colusão tácita ou estratégias instáveis — pode-se afirmar que ele foi **adequadamente respondido**. Os resultados revelaram que os agentes tendem a adotar estratégias que oscilam entre o equi-

líbrio de Cournot e padrões de colusão tácita, com clara predominância da coordenação emergente em ambientes mais populosos. Não foi observada instabilidade generalizada, tampouco comportamento caótico, o que reforça a robustez dos algoritmos de aprendizado utilizados e sua capacidade de gerar comportamentos economicamente plausíveis.

A metodologia adotada foi de natureza experimental-computacional, baseada em simulações conduzidas com modelagem baseada em agentes. O experimento foi implementado na plataforma NetLogo, utilizando a extensão `qlearningextension` para operacionalizar o algoritmo Q-Learning. As simulações foram automatizadas pelo módulo *BehaviourSpace*, o que permitiu a execução de dezenas de cenários com variações paramétricas sistemáticas. Os dados gerados foram exportados no formato `.csv` e analisados estatisticamente no ambiente R com suporte do pacote `tidyverse`. Essa abordagem metodológica garantiu alto grau de controle, replicabilidade e robustez estatística na análise dos padrões emergentes.

Apesar dos avanços e das contribuições obtidas, a pesquisa apresenta algumas limitações relevantes. Em primeiro lugar, a escolha exclusiva do algoritmo Q-Learning, embora justificada pela sua simplicidade e estabilidade, restringe a generalização dos resultados. Métodos alternativos de aprendizado por reforço — como SARSA, Deep Q-Networks (DQN) ou políticas baseadas em atores-críticos — poderiam oferecer dinâmicas de aprendizado distintas, especialmente em contextos com maior dimensionalidade ou complexidade estratégica.

Em segundo lugar, a configuração dos parâmetros ambientais foi deliberadamente simplificada, com custos marginais nulos e função de demanda linear. Embora isso tenha facilitado a comparação com benchmarks teóricos, limita a aproximação com mercados reais, nos quais estruturas de custo e elasticidades variam significativamente. A introdução de assimetrias entre firmas, choques exógenos ou dinâmicas de entrada e saída poderia enriquecer os cenários simulados.

Além disso, o espaço de ação dos agentes foi discretizado, e a memória limitada à atualização da matriz Q. Esses fatores podem ter restringido a capacidade dos agentes de desenvolver estratégias mais sofisticadas, como retaliações dinâmicas ou reconhecimento de padrões complexos. A ausência de comunicação entre agentes também exclui a possibilidade de investigar explicitamente mecanismos de conluio, como punições ou acordos implícitos com base em sinais públicos.

Dadas essas limitações, algumas recomendações podem ser feitas para pesquisas futuras. Em primeiro lugar, sugere-se a adoção de algoritmos mais avançados de aprendizado por reforço, capazes de capturar nuances estratégicas mais sutis. O uso de redes neurais profundas ou métodos bayesianos, por exemplo, poderia ampliar significativamente a sofisticação dos agentes e a diversidade de comportamentos possíveis.

Em segundo lugar, propõe-se a introdução de ambientes com maior heterogenei-

dade estrutural, incluindo firmas assimétricas, mudanças nas condições de mercado e choques de demanda. Essa abordagem permitiria testar a robustez dos resultados em contextos mais realistas e avaliar a resiliência das estratégias aprendidas frente à incerteza.

Por fim, futuras pesquisas poderiam explorar a interação entre diferentes tipos de agentes — por exemplo, combinando agentes racionais, imitadores e adaptativos — para investigar como diferentes heurísticas de decisão se propagam em mercados simulados. Tais estudos contribuiriam não apenas para o refinamento da teoria dos jogos evolucionários, mas também para aplicações práticas em regulação antitruste, modelagem de mercados digitais e simulações de políticas públicas.

Referências

TESFATSION, L. Agent-based computational economics: A constructive approach to economic theory. In: TESFATSION, L.; JUDD, K. L. (Ed.). **Handbook of Computational Economics, Volume 2**. [S.l.]: Elsevier, 2006. p. 831–880.

EPSTEIN, J. M. Agent-based computational models and generative social science. **Complexity**, v. 4, n. 5, p. 41–60, 1999.

AXELROD, R. **The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration**. [S.l.]: Princeton University Press, 1997.

ARTHUR, W. B. Inductive reasoning and bounded rationality. **The American Economic Review**, v. 84, n. 2, p. 406–411, 1994.

WATKINS, C. J.; DAYAN, P. Technical note: Q-learning. **Machine Learning**, Springer, v. 8, n. 3-4, p. 279–292, 1992. Disponível em: https://www.researchgate.net/publication/220344150_Technical_Note_Q-Learning.

XU, J. Reinforcement learning in a cournot oligopoly model. **Computational Economics**, v. 58, n. 4, p. 1001–1024, 2021. Disponível em: https://doi.org/10.1007/s10614-020-09982-4.

SHI, Y.; ZHANG, B. Multi-agent reinforcement learning in cournot games. In: **2020 59th IEEE Conference on Decision and Control (CDC)**. [S.l.: s.n.], 2020. p. 3561–3566. ABNT: SHI, Yuanyuan; ZHANG, Baosen. Multi-Agent Reinforcement Learning in Cournot Games. In: *IEEE Conference on Decision and Control (CDC)*, 59., 2020. Anais [...]. p. 3561–3566. DOI: https://doi.org/10.1109/CDC42340.2020.9304089.

LEBARON, B. Agent-based computational finance. **Handbook of Computational Economics**, Elsevier, v. 2, p. 1187–1233, 2006.

MYATT, D. P.; WALLACE, C. Cournot competition and the social value of information. **Journal of Economic Theory**, v. 158, p. 466–506, 2015. Disponível em: https://www.sciencedirect.com/science/article/pii/S0022053114001045.

LIAN, Z.; ZHENG, J. A dynamic model of cournot competition for an oligopolistic market. **Mathematics**, v. 9, n. 5, p. 489, 2021. Disponível em: https://www.mdpi.com/2227-7390/9/5/489.

WALTMAN, L.; KAYMAK, U. Q-learning agents in a cournot oligopoly model. **Journal of Economic Dynamics and Control**, Elsevier, v. 32, n. 10, p. 3275–3293, oct 2008. Disponível em: https://doi.org/10.1016/j.jedc.2008.01.003.

SCHELLING, T. C. Dynamic models of segregation. **Journal of Mathematical Sociology**, v. 1, n. 2, p. 143–186, 1971.

LEBARON, B. Agent-based computational finance: Suggested readings and early research. **Journal of Economic Dynamics and Control**, v. 24, n. 5–7, p. 679–702, 2000.

BRENNER, T. Agent learning representation: Advice on modelling economic learning. In: **Handbook of Computational Economics**. [S.l.]: Elsevier, 2006. v. 2, p. 895–947.

GILBERT, N.; TROITZSCH, K. **Simulation For The Social Scientist**. McGraw-Hill Education, 2005. ISBN 9780335216000. Disponível em: https://books.google.com.br/books?id=fBlaulpmNowC.

WILENSKY, U.; RAND, W. **An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo**. Cambridge: The MIT Press, 2015. Disponível em: http://www.jstor.org/stable/j.ctt17kk851. Acesso em: 22 mar. 2025. ISBN 9780262731898. Disponível em: http://www.jstor.org/stable/j.ctt17kk851.

BONABEAU, E. Agent-based modeling: Methods and techniques for simulating human systems. **Proceedings of the National Academy of Sciences**, v. 99, n. suppl_3, p. 7280–7287, 2002. Disponível em: https://www.pnas.org/doi/abs/10.1073/pnas.082080899.

DUFFY, J. Agent-based models and human subject experiments. In: TESFATSION, L.; JUDD, K. L. (Ed.). **Handbook of Computational Economics**. Elsevier, 2006. v. 2, cap. 19, p. 949–1011. Disponível em: https://ideas.repec.org/h/eee/hecchp/2-19.html.

PRISMA – Centro de Física Teórica e Computacional. **Tutorial de NetLogo**. 2010. https://cftc.ciencias.ulisboa.pt/PRISMA/capitulos/netlogo/topico3.php. Acesso em 22 de março de 2025.

RAILSBACK, S. F.; GRIMM, V. **Agent-Based and Individual-Based Modeling: A Practical Introduction**. 2. ed. Princeton, NJ: Princeton University Press, 2020. Acesso em 21 mar. 2025. ISBN 9780691190822. Disponível em: https://press.princeton.edu/books/hardcover/9780691190822/agent-based-and-individual-based-modeling.

BANOS, A.; LANG, C.; MARILLEAU, N.; CHéREL, G.; VOIRIN, T. **Agent-Based Spatial Simulation with NetLogo: Volume 1: Introduction and Bases**. London: Elsevier, 2015.

HORTON, N. J.; KLEINMAN, K. et al. Using r and rstudio for data management, statistical analysis, and graphics. CRC Press Boca Raton, FL, USA:, 2015.

KRONTHALER, F.; ZÖLLNER, S. Data analysis with rstudio. **Data Analysis with RStudio**, Springer, 2021.

BELLMAN, R. **Adaptive Control Processes: A Guided Tour**. Princeton: Princeton University Press, 1961. Disponível em: http://www.jstor.org/stable/j.ctt183ph6v.

LINS, R. A. S. **Aprendizagem por reforço profundo: uma nova perspectiva sobre o problema dos k-servos**. 2020. Tese (Tese (Doutorado em Engenharia Elétrica e de Computação)) — Universidade Federal do Rio Grande do Norte, Natal, Brasil, 2020. Disponível em: https://repositorio.ufrn.br/handle/123456789/29661.

BELLEMARE, M. G.; SRINIVASAN, S.; OSTROVSKI, G.; SCHAUL, T.; SAXTON, D.; MUNOS, R. Unifying count-based exploration and intrinsic motivation. **arXiv preprint arXiv:1606.01868**, June 2016. Disponível em: https://arxiv.org/abs/1606.01868. Acesso em: 20 mar. 2025. Disponível em: https://doi.org/10.48550/arXiv.1606.01868.

TANG, H.; HOUTHOOFT, R.; FOOTE, D.; STOOKE, A.; CHEN, X.; DUAN, Y.; SCHULMAN, J.; TURCK, F. D.; ABBEEL, P. #exploration : a study of count-based exploration for deep reinforcement learning. In: **Advances in Neural Information Processing Systems 30 (NeurIPS 2017)**. Long Beach, CA: [s.n.], 2017. p. 1–18. Disponível em: https://papers.nips.cc/paper_files/paper/2017/hash/2bd3811bcbd01c31c8692c071a68b7a5-Abstract.html. Acesso em: 20 mar. 2025. Disponível em: https://papers.nips.cc/paper_files/paper/2017/hash/2bd3811bcbd01c31c8692c071a68b7a5-Abstract.html.

SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning: An Introduction**. 2. ed. Cambridge, MA: The MIT Press, 2018. 552 p. Disponível em: https://mitpress.mit.edu/9780262039246. Acesso em: 20 mar. 2025. ISBN 9780262039246.

BELLMAN, R.; BELLMAN, R.; CORPORATION, R. **Dynamic Programming**. Princeton University Press, 1957. (Rand Corporation research study). Disponível em: https://books.google.com.br/books?id=rZW4ugAACAAJ.

DERMAN, C. **Finite State Markovian Decision Processes**. New York: Academic Press, 1970. ISBN 9780122092503. Disponível em: https://lib.ugent.be/catalog/ebk01:2550000000015205.

PUTERMAN, M. L. **Markov Decision Processes: Discrete Stochastic Dynamic Programming**. 1st. ed. USA: John Wiley & Sons, Inc., 1994. ISBN 0471619779.

BOYAN, J.; MOORE, A. Generalization in reinforcement learning: Safely approximating the value function. In: TESAURO, G.; TOURETZKY, D.; LEEN, T. (Ed.). **Advances in Neural Information Processing Systems**. MIT Press, 1994. v. 7. Disponível em: https://proceedings.neurips.cc/paper_files/paper/1994/file/ef50c335cca9f340bde656363ebd02fd-Paper.pdf.

SZEPESVáRI, C.; LITTMAN, M. L. A unified analysis of value-function-based reinforcement-learning algorithms. **Neural Computation**, v. 11, n. 8, p. 2017–2060, 11 1999. ISSN 0899-7667. Disponível em: https://doi.org/10.1162/089976699300016070.

LITTMAN, M. L. Value-function reinforcement learning in markov games. **Cognitive Systems Research**, v. 2, n. 1, p. 55–66, 2001. ISSN 1389-0417. Disponível em: https://www.sciencedirect.com/science/article/pii/S1389041701000158.

DAUGHETY, A. F. Cournot competition. In: DURLAUF, S. N.; BLUME, L. E. (Ed.). **The New Palgrave Dictionary of Economics**. Palgrave Macmillan, 2006. Forthcoming. Disponível em: https://cdn.vanderbilt.edu/vu-my/wp-content/uploads/sites/1683/2019/04/14130132/CournotCompetition-Daughety-webversion.pdf.

GRISáKOVá, N.; ŠTETKA, P. Cournot's oligopoly equilibrium under different expectations and differentiated production. **Games**, v. 13, n. 6, 2022. ISSN 2073-4336. ABNT: GRISÁKOVÁ, Nora; ŠTETKA, Peter. Cournot's Oligopoly Equilibrium under Different Expectations and Differentiated Production. *Games*,

v. 13, n. 6, art. 82, 2022. DOI: https://doi.org/10.3390/g13060082. Disponível em: https://www.mdpi.com/2073-4336/13/6/82.

ALóS-FERRER, C. Cournot versus walras in dynamic oligopolies with memory. **International Journal of Industrial Organization**, v. 22, n. 2, p. 193–217, 2004. ISSN 0167-7187. Disponível em: https://www.sciencedirect.com/science/article/pii/S0167718703001280.

FARMER, J. D.; FOLEY, D. The economy needs agent-based modelling. **Nature**, v. 460, n. 7256, p. 685–686, 2009.

RUST, J. Using randomization to break the curse of dimensionality. **Econometrica**, v. 65, n. 3, p. 487–516, 1997.

VARIAN, H. R. **Microeconomic Analysis**. 3rd. ed. [S.l.]: W.W. Norton & Company, 1992. ISBN 9780393957358.

CÓDIGOS

A.1 CÓDIGO NETLOGO

```
extensions [qlearningextension]

breed [empresas empresa]
empresas-own [quantidade lucro quantidade-total ]

globals [P Q capacidade-maxima]

to setup
  clear-all

  ;; Configura o inicial
  set capacidade-maxima 20

  create-empresas num-empresas [
    set quantidade random 20 + 1
    set quantidade-total 0
    set lucro 0
  ]

  ;; Configura o do Q-Learning para as empresas
  ask empresas [
    qlearningextension:state-def ["quantidade"]

    ;; Lista das a es dispon veis para o Q-learning
    (qlearningextension:actions [es_1] [es_2] [es_3] [es_4] [es_5] [
       es_6] [es_7] [es_8] [es_9] [es_10] [es_11] [es_12] [es_13] [
       es_14] [es_15] [es_16] [es_17] [es_18] [es_19] [es_20])

    ;; Outras configura es do Q-learning
    qlearningextension:reward [lucroFunc]
    qlearningextension:end-episode [isEndState] resetEpisode
    qlearningextension:action-selection "e-greedy" (list FE FDM)
```

```
    qlearningextension:learning-rate 0.005
    qlearningextension:discount-factor 0.9
  ]

  reset-ticks
end

to go
  ask empresas [
    qlearningextension:learning
  ]

  ;; Atualiza o valor de Q e calcula o pre o e lucro das empresas
  set Q sum [quantidade] of empresas
   set P A - Q


  ask empresas [
    set lucro P * quantidade
    set quantidade-total quantidade-total + quantidade
  ]


  tick
end

;; A   es poss veis com restri  o expl cita

to es_1 set quantidade 1 end
to es_2 set quantidade 2 end
to es_3 set quantidade 3 end
to es_4 set quantidade 4 end
to es_5 set quantidade 5 end
to es_6 set quantidade 6 end
to es_7 set quantidade 7 end
to es_8 set quantidade 8 end
to es_9 set quantidade 9 end
to es_10 set quantidade 10 end
to es_11 set quantidade 11 end
to es_12 set quantidade 12 end
to es_13 set quantidade 13 end
to es_14 set quantidade 14 end
```

```
to es_15 set quantidade 15 end
to es_16 set quantidade 16 end
to es_17 set quantidade 17 end
to es_18 set quantidade 18 end
to es_19 set quantidade 19 end
to es_20 set quantidade 20 end


;; Fun  o de recompensa
to-report lucroFunc
  let minha-q quantidade
  let Q-alheia sum [quantidade] of other empresas
  let P-local A - (Q-alheia + minha-q)
  report P-local * minha-q
end


;; Estado final do epis dio (encerra ap s 500 ticks)
to-report isEndState
  if ticks > 500 [report true]
  report false
end

;; Reinicia as vari veis para um novo epis dio
to resetEpisode
  set quantidade-total 0
  set lucro 0
```

A.2 CÓDIGO R

```
#      Instalar pacotes necess rios (execute apenas uma vez)
install.packages('readxl')
install.packages("tidyverse")
install.packages('janitor')
install.packages("sandwich")
install.packages("lmtest")

#      Carregar pacotes
library(readxl)    # Para ler arquivos Excel
library(tidyverse) # Inclui dplyr e ggplot2 para manipula  o e
   visualiza  o de dados
```

```r
library(dplyr)     # Para opera  es em dataframes
library(janitor)  # Para limpar e padronizar nomes de colunas

#      Definir o caminho do arquivo Excel contendo os dados
rota_dedados <- "C:\\Users\\xaide\\OneDrive\\ rea  de Trabalho\\Dados
   \\Novos Dados.xlsx"

#      Listar as planilhas dispon veis no arquivo (opcional, para
   confer ncia)
excel_sheets(rota_dedados)

#      Importar os dados da simula  o (ajuste o range conforme
   necess rio)
aprendizado_do_modelo <- read_excel(rota_dedados, range = 'A2:G1622')
    %>%
  clean_names() # Torna os nomes das colunas mais f ceis de usar (
     sem espa os, mai sculas etc.)

#         Visualizar os dados
head(aprendizado_do_modelo)

#      Garantir que os dados est o no formato certo
reg_data <- aprendizado_do_modelo %>%
  mutate(
    fe = as.numeric(fe),
    fdm = as.numeric(fdm),
    a = as.numeric(a),
    num_empresas = as.numeric(num_empresas),
    mean_lucro_of_empresas = as.numeric(mean_lucro_of_empresas)
  ) %>%
  drop_na(mean_lucro_of_empresas)

#      Modelo 1 - Regress o Linear Simples
modelo_lucro <- lm(mean_lucro_of_empresas ~ fe + fdm + a + num_
   empresas, data = reg_data)
summary(modelo_lucro)

#      Diagn stico dos res duos
par(mfrow = c(2, 2))  # 4 gr ficos em 1 janela
plot(modelo_lucro)
par(mfrow = c(1, 1))  # voltar ao normal

```

```r
#      Modelo 2 - Regress o com intera  o (fe * num_empresas)
modelo_interacao <- lm(mean_lucro_of_empresas ~ fe * num_empresas +
    fdm + a, data = reg_data)
summary(modelo_interacao)

#      Modelo 3 - Regress o com log do lucro
modelo_log <- lm(log(mean_lucro_of_empresas + 1) ~ fe + fdm + a + num
    _empresas, data = reg_data)
summary(modelo_log)

#      Modelo 4 - Regress o robusta com erros padr o robustos

library(sandwich)
library(lmtest)

# Coeficientes com erro robusto (modelo linear simples)
coeftest(modelo_lucro, vcov = vcovHC(modelo_lucro, type = "HC1"))
coeftest(modelo_interacao, vcov = vcovHC(modelo_interacao, type = "
    HC1"))
coeftest(modelo_log, vcov = vcovHC(modelo_log, type = "HC1"))

#         Ver os dados
View(comparacao_lucros)

#      Visualiza  o da compara  o
library(ggplot2)
comparacao_long <- comparacao_lucros %>%
  pivot_longer(cols = starts_with("lucro"), names_to = "modelo",
      values_to = "lucro")

ggplot(comparacao_long, aes(x = as.factor(num_empresas), y = lucro,
    fill = modelo)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ a) +
  labs(
    title = "Compara  o dos Lucros M dios: Q-Learning vs Cournot
        vs Colusivo",
    x = "N mero de Empresas",
    y = "Lucro M dio",
    fill = "Modelo"
  ) +
  theme_minimal()
```

```r

#      Ver lucro m dio por firma variando A e n mero de empresas
lucros_por_grupo <- reg_data %>%
  group_by(num_empresas, a) %>%
  summarise(
    lucro_medio = mean(mean_lucro_of_empresas, na.rm = TRUE)
  ) %>%
  arrange(num_empresas, a)

#         Ver os dados em formato de tabela
View(lucros_por_grupo)

# Adicionar lucros te ricos (Nash e Colus o)
lucros_por_grupo <- lucros_por_grupo %>%
  mutate(
    lucro_nash_teorico = (a^2) / (num_empresas + 1)^2,
    lucro_colusao_teorico = (a^2) / (4 * num_empresas)
  )

view(lucros_por_grupo)


lucros_por_grupo <- lucros_por_grupo %>%
  mutate(
    desvio_pct_nash = 100 * (lucro_medio - lucro_nash_teorico) /
       lucro_nash_teorico,
    desvio_pct_colusao = 100 * (lucro_medio - lucro_colusao_teorico)
       / lucro_colusao_teorico
  )

# Ver os desvios em tabela
View(lucros_por_grupo)

lucros_long <- lucros_por_grupo %>%
  pivot_longer(cols = c("lucro_medio", "lucro_nash_teorico", "lucro_
     colusao_teorico"),
               names_to = "modelo", values_to = "lucro")

ggplot(lucros_long, aes(x = as.factor(num_empresas), y = lucro, fill
   = modelo)) +
  geom_bar(stat = "identity", position = "dodge") +
  facet_wrap(~ a, scales = "free_y") +
```

```
121   labs(
122     title = "Lucro por Firma: Q-Learning vs Nash vs Colus o",
123     x = "N mero de Empresas",
124     y = "Lucro",
125     fill = "Modelo"
126   ) +
127   theme_minimal()
128
129
130 desvios_long <- lucros_por_grupo %>%
131   pivot_longer(cols = starts_with("desvio_pct"), names_to = "
        comparacao", values_to = "desvio_pct")
132
133 ggplot(desvios_long, aes(x = as.factor(num_empresas), y = desvio_pct,
        fill = comparacao)) +
134   geom_bar(stat = "identity", position = "dodge") +
135   facet_wrap(~ a) +
136   labs(
137     title = "Desvio Percentual do Q-Learning em rela  o a Nash e
          Colus o",
138     x = "N mero de Empresas",
139     y = "Desvio (%)",
140     fill = "Compara  o"
141   ) +
142   theme_minimal()
```