UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
ESCUELA DE POSTGRADO Y EDUCACIÓN CONTINUA

IMPACTO DE INFLACIÓN DE PRECIOS EN LA COLUSIÓN ALGORÍTMICA PARA PRICING

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN CIENCIA DE DATOS

SEBASTIÁN MARTÍN TINOCO PÉREZ

PROFESOR GUÍA:
ANDRÉS ABELIUK KIMELMAN

MIEMBROS DE LA COMISIÓN:
JAVIER RUIZ DEL SOLAR
MARCELO OLIVARES ARENAS

Este trabajo ha sido patrocinado por:
National Center for Artificial Intelligence CENIA FB210017, Basal ANID

SANTIAGO DE CHILE
2024

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE MAGÍSTER EN CIENCIAS
DE DATOS
POR: SEBASTIÁN MARTÍN TINOCO PÉREZ
FECHA: 2024
PROF. GUÍA: ANDRÉS ABELIUK KIMELMAN

IMPACTO DE INFLACIÓN DE PRECIOS EN LA COLUSIÓN ALGORÍTMICA PARA PRICING

En un contexto social marcado por el auge de algoritmos autónomos y elevadas tasas de inflación, resulta imperativo evaluar los potenciales riesgos que estos algoritmos pueden representar para la competitividad de los mercados. Esta tesis explora la cooperación emergente en el pricing algorítmico bajo condiciones de costos no estacionarios, utilizando algoritmos de aprendizaje reforzado. Se investigó cómo los agentes económicos ajustan sus políticas de precios en respuesta a variaciones en los costos de producción y shocks inflacionarios. Los resultados indican que la inflación tiene un impacto significativo en las rentabilidades de los agentes, sugiriendo una disminución en la competitividad del mercado. Además, se evaluaron estrategias de castigo y la sensibilidad a hiperparámetros tanto del modelo económico como de los algoritmos, destacando la importancia de una correcta configuración de estos para mantener la estabilidad del mercado. La inclusión de un agente altruista mostró ser efectiva para aumentar la competitividad, aunque presenta desafíos en su implementación. Esta tesis subraya la necesidad de diseñar políticas y regulaciones que aseguren la equidad en mercados cada vez más influenciados por la inteligencia artificial.

A Macarena y Colihue,
por su amor y apoyo incondicional.

Gracias.

Agradecimientos

Reflexionando sobre mi experiencia en estos últimos años, jamás habría imaginado el nivel de aprendizaje y crecimiento que obtendría a través de este programa de Magíster. Quisiera dedicar este momento a reconocer a todas aquellas personas que, de un modo u otro, fueron esenciales en mi proceso educativo y dejaron una huella imborrable en mi paso por la facultad de Beauchef.

Primero, quiero expresar mi más profundo agradecimiento a la comisión del Magíster en Ciencia de Datos por otorgarme la oportunidad de ser parte de este desafiante programa. Viniendo de un ámbito como la Ingeniería Comercial, tenía serias reservas sobre si contaba con los conocimientos matemáticos y computacionales requeridos. Sin embargo, gracias a esta oportunidad, pude enfrentar y superar varios desafíos. Un agradecimiento especial a Mauricio Bernier, administrativo del Magíster, cuya dedicación y disposición para resolver dudas y problemas fueron excepcionales.

En segundo lugar, quiero reconocer a mis profesores guía, cuya influencia fue crucial en el desarrollo de mi tesis. Agradezco especialmente al profesor Andrés Abeliuk por su gran disponibilidad y por permitirme explorar un tema de mi elección. Al profesor Javier Ruiz del Solar, agradezco su apoyo en los aspectos técnicos y algorítmicos de mi trabajo, siendo un pilar fundamental para la validez de mis resultados. Y al profesor Marcelo Olivares, por su contribución en la discusión y formulación del modelo económico y el diseño experimental.

En tercer lugar, me gustaría agradecer al equipo de data AB InBev, en especial a Lauta, Benja y Emi. Gracias por confiar en mi y permitirme desarrollar el trabajo de tesis al mismo tiempo que el trabajo. Esa confianza es importante pues me dió las herramientas para avanzar y cerrar finalmente este proceso. Sepan que si bien nos dejaremos de ver en el dia a dia, fueron parte fundamental de mi paso por la compañía y el cierre de este proceso también se los debo a ustedes. Siempre los recordaré con mucho cariño.

En cuarto lugar, deseo agradecer de corazón a todos los amigos y compañeros que conocí durante el programa de Magíster. En particular, quiero agradecer de forma especial a Ignacio Meza, Stefano Schiappacase y Samuel Molina, quienes no solo compartieron conmigo innumerables momentos de risa y compañerismo, sino que también fueron un apoyo constante y me motivaron a buscar la excelencia en cada entrega. Sepan que mi experiencia con este magíster no habría sido la misma sin ustedes y les estaré por siempre agradecido por eso.

De forma adicional, quiero agradecer a National Center for Artificial Intelligence CENIA FB210017, Basal ANID por su apoyo en el desarrollo de esta tesis.

Finalmente, deseo extender mi más profundo y especial agradecimiento a mi familia y seres queridos, quienes fueron esenciales para completar este Magíster. A mi madre, Marta, por su constante preocupación y apoyo incondicional. A mi hermana, Natalia, por su amor, alegría, y optimismo, que siempre iluminaron mi camino. A mis queridos chucaos: a Colihue, mi fiel compañero perruno, por traer alegría y ternura a nuestras vidas (aunque a veces sea un poco *stubborn*), y, en especial, a Macarena, por su amor, apoyo, y paciencia a lo largo de todo este proceso. Sin duda, has sido la parte más importante de esta etapa, y sin ti, nada de esto habría sido posible.

A todos y cada uno, y desde el fondo de mi corazón: Muchas gracias.

Tabla de Contenido

Índice de Tablas

Índice de Ilustraciones

Capítulo 1

Introducción

1.1. Problema

En la era contemporánea, el comercio electrónico ha registrado un crecimiento sin precedentes, marcado por un aumento exponencial en las compras en línea. Como señala Tudor [1], este auge, acelerado por la pandemia de COVID-19, ha llevado a que las ventas de retail en Estados Unidos a través de plataformas digitales superen los 227 billones de dólares, reflejando un cambio significativo en los hábitos de consumo hacia la digitalización. En este contexto, las empresas, siguiendo el ejemplo de líderes del sector como Amazon, están adoptando cada vez más algoritmos de inteligencia artificial para decisiones críticas, en particular aquellos basados en el Aprendizaje Reforzado. Este enfoque busca maximizar los beneficios a través de una estrategia de precios dinámica, que se adapta a las variaciones del mercado, como cambios en la oferta y las tácticas de precios competitivos implementadas por otros jugadores del mercado.

No obstante, la transición hacia la automatización de decisiones esenciales en el ámbito empresarial conlleva consecuencias inadvertidas, particularmente en términos de competitividad de mercado. Investigaciones recientes, enmarcadas en la literatura de *Algorithmic Collusion*, como las de Calvano et al. [2] y Klein et al. [3], han revelado que los algoritmos de aprendizaje por refuerzo, incluso en configuraciones relativamente sencillas, tienen la capacidad de aprender a coludirse. Esta colusión se manifiesta en la fijación de precios por encima del equilibrio competitivo de Nash, sin necesidad de comunicación explícita entre las empresas. Esta forma de colusión tácita, potenciada por el historial de precios y la capacidad de los algoritmos para penalizar desviaciones, plantea un significativo desafío para la eficiencia del mercado y podría culminar en un escenario donde las corporaciones se benefician a expensas de los consumidores.

En un contexto económico marcado por elevadas tasas de inflación, resultado de políticas monetarias agresivas implementadas para mitigar los efectos de la pandemia de COVID-19, resulta imperativo evaluar si estos algoritmos mantendrán su tendencia a establecer precios supra-competitivos o si adaptarán sus estrategias de una manera que podría intensificar o mitigar los efectos inflacionarios en los precios al consumidor. Esta línea de investigación es crucial no solo para entender mejor las dinámicas de mercado en periodos de inestabilidad económica, sino también para el diseño de políticas y regulaciones efectivas que aseguren un equilibrio y justicia en un mercado cada vez más influenciado por la inteligencia artificial.

1.2. Hipótesis

Considerando la creciente adopción de algoritmos de Aprendizaje Reforzado por parte de las empresas para la toma de decisiones críticas y las elevadas tasas de inflación observadas en los últimos años, esta investigación plantea la hipótesis de que dichos algoritmos, debido a su notable capacidad de adaptación, ajustan sus políticas de precios frente a shocks inflacionarios y variaciones en los costos de producción de manera que mantienen o incrementan precios supra-competitivos, afectando negativamente la competitividad de los mercados.

1.3. Objetivos

1.3.1. Objetivo general

Demostrar que la competencia de modelos de pricing basados en aprendizaje reforzado conlleva a comportamientos emergentes de cooperación, y que estos persisten aún cuando los agentes se enfrentan a shocks inflacionarios en los costos de producción.

1.3.2. Objetivos específicos

Se proponen los siguientes objetivos específicos:

- Diseñar un ambiente experimental basado en la literatura *Algorithmic Collusion* en el que se implementen costos de producción variables.
- Cuantificar el grado de cooperación de agentes y su reacción ante shocks en los costos.
- Evaluar el impacto de los hiperparámetros tanto del modelo económico como de los algoritmo en los resultados obtenidos
- Demostrar la existencia de estrategias de castigo emergentes ante desvíos del equilibrio

1.4. Contribución a la Literatura

Este trabajo realiza una serie de aportes significativos a la literatura existente en el ámbito del pricing algorítmico y el aprendizaje reforzado, destacándose en los siguientes aspectos:

- **Formulación del problema con costos variables**: Se formula el problema incorporando la variabilidad en los costos de producción mediante la inclusión de shocks inflacionarios. Esta aproximación permite una evaluación más realista de las estrategias de pricing en contextos económicos dinámicos.
- **Comparación exhaustiva de resultados**: Se realiza una comparación detallada de los resultados obtenidos en ambientes de entrenamiento y prueba, proporcionando una visión integral sobre la robustez y generalización de las estrategias aprendidas por los agentes.
- **Desafío a los supuestos tradicionales**: Se cuestiona y levanta el supuesto tradicional de que los equilibrios de Nash y Monopolio son conocidos a priori por los agentes. En su lugar, se analiza cómo los agentes pueden aprender y adaptarse a estos equilibrios en un entorno donde la información es imperfecta y los costos son no estacionarios.

1.5. Estructura de la Tesis

Esta tesis se estructura en cinco capítulos, cada uno proporcionando una visión preliminar de los temas abordados. El Capítulo 2, "Marco Teórico", establece los fundamentos conceptuales, abarcando aspectos clave tanto del ámbito económico como del aprendizaje reforzado. El Capítulo 3, "Metodología", describe en detalle las técnicas empleadas y el diseño experimental del estudio. El Capítulo 4 se dedica a la presentación y análisis de los resultados obtenidos. Por último, el Capítulo 5 concluye el trabajo, resumiendo los hallazgos principales y esbozando direcciones para investigaciones futuras.

Capítulo 2

Marco Teórico

2.1. Modelo Económico

2.1.1. Competencia en Oligopolio

La competencia en oligopolio constituye un fenómeno central en el ámbito de la microeconomía, desempeñando un papel esencial en la comprensión de las dinámicas del mercado y la toma de decisiones estratégicas por parte de las empresas. Este tipo de competencia se caracteriza por la presencia de un reducido número de empresas que dominan el mercado, y cuyas acciones y estrategias tienen un impacto significativo en la determinación de precios, la oferta y la demanda de bienes y servicios.

En el oligopolio, las empresas se encuentran interdependientes, lo que significa que las decisiones de una empresa afectan directamente las estrategias y el desempeño de las demás. Este contexto de interacción estratégica lleva a situaciones donde las empresas deben considerar cuidadosamente cómo ajustar sus precios, cantidades producidas y estrategias publicitarias para maximizar sus propios beneficios en un entorno competitivo. Como consecuencia de esto, es usual que los estudios de mercado efectuados para este tipo de competencia se realicen desde la disciplina de la *Teoría de Juegos*.

Desde la perspectiva de la microeconomía, el oligopolio ofrece un terreno fértil para el análisis detallado de conceptos como la maximización de beneficios, la colusión entre empresas, la competencia no cooperativa, y la formación y mantenimiento de barreras de entrada. Los modelos y teorías desarrolladas en este contexto permiten a los economistas y analistas comprender las complejidades de la toma de decisiones empresariales en mercados oligopólicos.

La importancia de estudiar la competencia en oligopolio radica en su influencia directa en la competitividad de los mercados. Las acciones estratégicas de las empresas en este tipo de estructura de mercado pueden tener impactos significativos en los precios y la calidad de los productos, así como en la innovación y la eficiencia del mercado en general. Por lo tanto, la comprensión de las dinámicas del oligopolio es esencial para evaluar y fomentar la competitividad efectiva en diversos sectores económicos.

2.1.2. Competencia de Bertrand: Modelo clásico

La Competencia de Bertrand [4], denominada en honor a Joseph Louis François Bertrand, constituye un modelo económico de competencia oligopólica en el que empresas venden sus productos a través de la fijación de precios. A su vez, los consumidores eligen la cantidad a adquirir en función de los precios percibidos. A diferencia de otros modelos económicos y como resultado de la competencia en precios, el modelo de Bertrand propone que el precio de equilibrio competitivo es equivalente al costo marginal de producción, recuperando de esta manera el equilibrio obtenido en un mercado con competencia perfecta.

En su versión clásica, el modelo ecónomico de Bertrand posee los siguientes supuestos:

- Existen al menos 2 firmas en el mercado
- Cada firma i busca maximizar sus beneficios $\mathbb{R}_i$
- Las firmas compiten solo 1 período
- Los productos de cada empresa son homogéneos e indiferenciados
- Las firmas tienen un costo marginal de producción c fijo e idéntico
- Ambas firmas escogen y revelan sus precios p_i de manera simultánea
- Consumidores compran q_i unidades a la firma con menor precio

Considerando lo anterior, el problema de optimización viene determinado por:

$$\max_{p_i} R_i = (p_i - c) \cdot q_i \tag{2.1}$$

De igual manera, un aspecto usual para la comunidad académica es la reformulación de este modelo a T períodos:

$$\max_{p_{i,t}} R_i = \sum_t (p_{i,t} - c) \cdot q_{i,t} \tag{2.2}$$

donde t indica el período $t \in T$. A partir de la formulación anterior, es interesante analizar los equilibrios de mercado resultantes al variar el nivel de cooperación entre las empresas.

2.1.3. Caso Libre Competencia: Equilibrio de Nash

En el caso donde las firmas participantes compitan de forma individual y considerando que los consumidores prefieren comprar a la firma con un menor precio, la estrategia óptima de la firma i será siempre vender sus productos a un precio marginalmente más bajo que su competidor j siempre que este nuevo precio esté encima de sus costos c, es decir:

$$p_{i,t} = \begin{cases} p_{j,t} - \epsilon & \text{si } p_{j,t} - \epsilon > c \\ c & \text{en otro caso} \end{cases} \tag{2.3}$$

La Figura 2.1 presenta la *función de reacción* de cada firma como respuesta al precio fijado por su competencia:

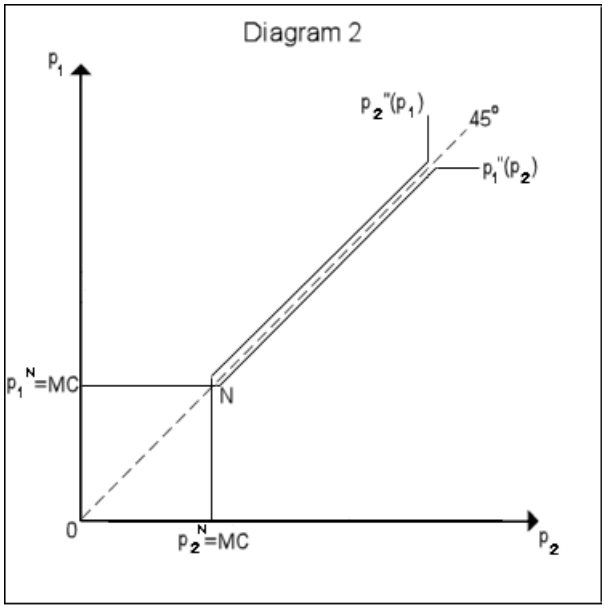


Figura 2.1: Función de reacción de empresas en Competencia a la Bertrand [5].

Es importante notar que al momento de fijar los precios del periodo t, **la firma i no conoce el precio fijado por su competencia** $p_{j,t}$, por lo que teóricamente solo puede *especular* el precio fijado por la competencia y esperar que el precio fijado sea efectivamente menor al efectuado por esta empresa. Lo mismo pasa al revés: la estrategia óptima de la empresa j es fijar un precio marginalmente inferior a la empresa i, pero al momento de fijar su precio esta desconoce $p_{i,t}$ y para esto solo puede esperar que su precio sea menor al de la firma i. Finalmente, ambas empresas conocen que su competencia incorpora sus propias acciones para fijar su precio, es decir:

$$p_{i,t} = \begin{cases} p_{j,t}(p_{i,t}) - \epsilon & \text{si } p_{j,t} - \epsilon > c \\ c & \text{en otro caso} \end{cases} \tag{2.4}$$

donde el término $p_{j,t}(p_{i,t})$ representa el precio fijado por la firma j tomando en cuenta el precio fijado por la firma i.

Este tipo de problemática donde el resultado *depende* de las acciones de los otros jugadores es algo usual en el campo de *Teoría de Juegos* y da lugar a la generación de diferentes estrategias para maximizar el beneficio obtenido por el jugador, las que a su vez generan diferentes *equilibrios* como resultado. Analizando la Figura 2.1 es fácil ver que, para cualquier resultado donde el precio p_j sobre c, siempre será óptimo fijar p_i marginalmente inferior a p_j. La existencia de incentivos a desviarse de este equilibrio le otorga el nombre de *Equilibrio Débil*, pues es un resultado que fácilmente puede tener desestabilizarse y obtener fluctuaciones. Por otro lado, cuando p_j es igual a c, ya no es óptimo fijar un precio inferior a p_j, pues de hacerlo se incurriría en un rentabilidades negativas. Dado que en este punto no existen incentivos unilaterales a deviarse de este equilibrio, se concluye que $p_i = p_j = c$ es un *Equilibrio de Nash en Estrategias Mixtas*:

$$p^N = c \tag{2.5}$$

Donde tomando la ecuación 2.2, se desprende que:

$$R^N = 0 \tag{2.6}$$

La identidad obtenida en la ecuación 2.6 es conocida como la **Paradoja de Bertrand**, pues bajo esta configuración un duopolio es capaz de llegar al mismo equilibrio que el equilibrio de Competencia Perfecta (rentabilidades iguales a cero), desafiando la intuición de que un oligopolio puede establecer precios ligeramente por encima del costo marginal para obtener ganancias.

2.1.4. Diferenciación de productos

Una de las alternativas para sortear la Paradoja de Bertrand es a través del a inclusión de diferenciación en los productos transados, es decir, levantar el supuesto de que los productos sean homogéneos e indiferenciados. Una forma de representar esto último es a través de una función de demanda *logit* en su versión canónica, es decir:

$$q_{i,t} = \frac{e^{\frac{\alpha_i - p_{i,t}}{\mu}}}{\sum_{j=1}^{n} e^{\frac{\alpha_j - p_{j,t}}{\mu_i}} + e^{\frac{\alpha_0}{\mu}}} \tag{2.7}$$

donde

- $q_{i,t}$ corresponde a la cantidad recibida por la empresa i en el período t
- $p_{i,t}$ corresponde al precio fijado por la empresa i en el periodo t
- α_i corresponde a la diferenciación vertical entre los productos de las empresas
- μ corresponde a la diferenciación horizontal entre los productos de las empresas
- α_0 es el índice inverso de la demanda agregada

Este modelo, tanto en su versión logit como en formulaciones similares, ha sido ampliamente adoptado en la literatura de Microeconomía y en diversos estudios de mercado, demostrando su versatilidad para ajustarse a múltiples industrias (e.g. [6], [7], [8], [9], [10]). Además, es crucial destacar que, como en toda función de demanda, existe una relación inversa entre la cantidad demandada y el precio establecido por las firmas.

2.1.5. Diferenciación de productos: Equilibrio de Nash

Para obtener el equilibrio de Nash con productos diferenciados, es necesario encontrar las funciones de reacción de los agentes. Estas pueden ser encontradas diferenciando la función de beneficios tomando como referencia la función de demanda de la ecuación 2.7, es decir:

$$\frac{\partial R}{\partial p_i} = \frac{\partial}{\partial p_i}(p_i - c_t) \cdot q_{i,t}$$

$$\frac{\partial R}{\partial p_i} = \frac{\partial}{\partial p_i}(p_i - c_t) \cdot \frac{e^{\frac{\alpha_i - p_{i,t}}{\mu}}}{\sum_{j=1}^{n} e^{\frac{\alpha_j - p_{j,t}}{\mu_i}} + e^{\frac{\alpha_0}{\mu}}}$$

$$0 = \frac{e^{\frac{\alpha_i - p_{i,t}}{\mu}}}{e^{\frac{\alpha_0}{\mu}} + \sum_{j=1}^{n} e^{\frac{\alpha_j - p_{j,t}}{\mu}}} \left(1 + \frac{e^{\frac{\alpha_i - p_{i,t}}{\mu}} (-c_i + p_{i,t})}{\mu \left(e^{\frac{\alpha_0}{\mu}} + \sum_{j=1}^{n} e^{\frac{\alpha_j - p_{j,t}}{\mu}} \right)} - \frac{-c_i + p_{i,t}}{\mu} \right) \tag{2.8}$$

A partir de ambas funciones de reacción, el equilibrio de Nash puede ser encontrado igualando ambas funciones de reacción y encontrando la solución de forma simultánea para ambas identidades:

$$\frac{\partial R}{\partial p_i} - \frac{\partial R}{\partial p_j} = 0 \tag{2.9}$$

En la práctica y considerando la complejidad matemática para encontrar la expresión de la solución, es necesario recurrir a *softwares de optimización* y de esta manera obtener una solución aproximada al problema.

2.1.6. Caso Cooperativo: Equilibrio Monopólico

Considerando el caso donde existe colusión entre las empresas, es posible formular el problema de maximización de beneficios como un problema donde existe solo 1 empresa participando del mercado, es decir, un monopolio. De esta forma, la ecuación 2.7 se transforma a:

$$q_{i,t} = \frac{e^{\frac{\alpha_i - p_{i,t}}{\mu}}}{e^{\frac{\alpha_i - p_{i,t}}{\mu}} + e^{\frac{\alpha_0}{\mu}}} \tag{2.10}$$

Al mismo tiempo, el equilibrio monopólico se encuentra maximizando las rentabilidades en función del precio, es decir:

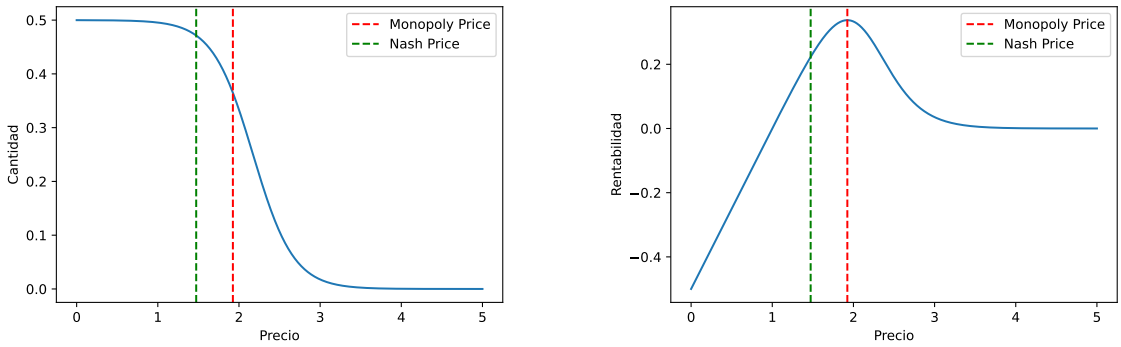
$$\max_{p_{i,t}} R_i = \sum_t (p_{i,t} - c_t) \cdot q_{i,t}$$

$$\max_{p_{i,t}} R_i = \sum_t (p_{i,t} - c_t) \cdot \frac{e^{\frac{\alpha_i - p_{i,t}}{\mu}}}{e^{\frac{\alpha_i - p_{i,t}}{\mu}} + e^{\frac{\alpha_0}{\mu}}}$$

donde nuevamente es usual recurrir a *softwares de optimización* para resolver el problema de optimización.

2.1.7. Análisis de rentabilidades

Habiendo determinado el equilibrio de Nash y de Monopolio, es interesante generar una comparación entre las rentabilidades teóricas obtenidas por ambos equilibrios. La Figura 2.2 presenta la comparativa entre ambos equilibrios, tanto a nivel de demanda como en términos de rentabilidades:



(a) Función de Demanda

(b) Curva de Rentabilidad

Figura 2.2: Análisis de Demanda y Rentabilidad.

A partir del gráfico anterior, se desprende que:

i) El Equilibrio de Monopolio p^M genera **rentabilidades estrictamente mayores** a las rentabilidades obtenidas por el Equilibrio de Nash p^N.

ii) El Equilibrio de Monopolio p^M posee un **nivel de precios estrictamente mayor** al nivel de precios obtenido por el Equilibrio de Nash p^N.

iii) Como consecuencia de un nivel de precios mayor, el equilibrio de Monopolio p^M tiene un **menor nivel de cantidad transada** que el equilibrio de Nash p^N.

De esta manera, el equilibrio de Monopolio es un resultado que es socialmente **no deseable** debido a que genera pérdidas de eficiencia, pues se obtiene a un equilibrio donde se venden menos unidades a un mayor precio que las que se podrían obtener en un contexto de libre competencia.

2.1.8. Propuesta: Evolución en los costos

2.1.8.1. Motivación

En la última década, la persistente presencia de la inflación ha adquirido una creciente relevancia en la formulación de estrategias de fijación de precios en diversos sectores económicos. La variabilidad en los niveles de precios, impulsada por las fluctuaciones inflacionarias, se presenta como un componente crítico que incide directamente en las decisiones de pricing adoptadas por empresas y agentes económicos. Esta relevancia se intensifica al considerar que el nivel de precios ha experimentado un notable aumento, triplicando su tasa de crecimiento en los últimos años en países desarrollados. Este fenómeno es atribuible a políticas monetarias agresivas implementadas como método de protección ante la pandemia de COVID-19, lo cual enfatiza aún más la importancia de una cuidadosa consideración de la inflación en las estrategias comerciales contemporáneas.

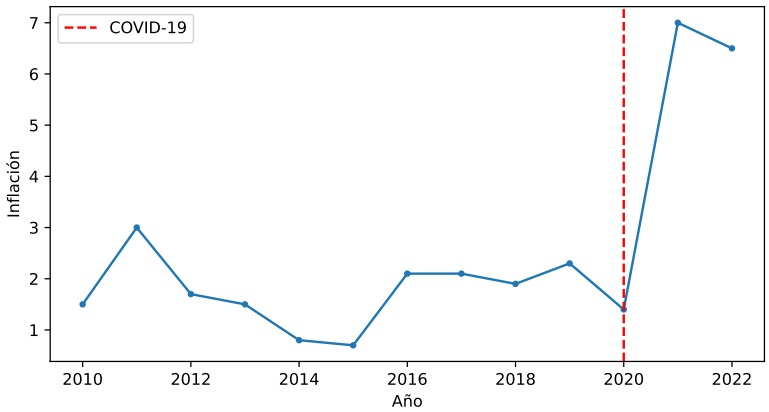


Figura 2.3: Tasa de Inflación Anual de Estados Unidos 2010-2022.

La capacidad de adaptación a entornos económicos dinámicos, caracterizados por cambios en las tasas de inflación, se ha consolidado como un factor determinante para el éxito y la sostenibilidad de las estrategias comerciales. En este contexto, la comprensión y consideración de la inflación como un componente integral en las decisiones de *pricing* se erige como un imperativo estratégico, subrayando la necesidad de investigaciones que exploren cómo los algoritmos empleados en la fijación de precios pueden responder eficazmente a las dinámicas inflacionarias contemporáneas.

Simultáneamente, junto con la incorporación del crecimiento en los costos de producción, es crucial modelar la dinámica del aumento en los niveles de precios desde la perspectiva de la demanda de los consumidores. Este enfoque es esencial para garantizar la perdurabilidad de los mercados ante incrementos en los niveles de precios. Un ejemplo ilustrativo de este escenario se materializa al considerar el precio de una caja de leche hace dos décadas: desembolsar tres dólares por dicha adquisición podría haberse percibido como extravagante en ese entonces, no obstante, en el actual contexto de precios, tal erogación se erige como una práctica común y aceptada.

La inclusión del crecimiento en los costos de producción y la disposición a pagar de los consumidores en la literatura de *Algorithmic Collusion* responde a la necesidad de abordar de manera sistemática y comprehensiva las implicaciones de las condiciones económicas cambiantes en el diseño y ejecución de estrategias algorítmicas por parte de los agentes económicos. La inflación, como indicador del aumento generalizado de precios en una economía, emerge como un elemento de significativa relevancia que puede ejercer influencia sobre la eficacia y estabilidad de prácticas colusivas implementadas mediante algoritmos. Esta consideración introduce una dimensión adicional de complejidad en el modelamiento y evaluación de las estrategias algorítmicas, demandando una adaptación dinámica por parte de los agentes para preservar o ajustar su política de precios en respuesta a las variaciones económicas. De esta manera, la inclusión del factor inflacionario en la investigación sobre *Algorithmic Collusion* persigue proporcionar un marco analítico más realista y robusto, capaz de reflejar las condiciones económicas del entorno y, por ende, facilitar una comprensión más integral de las interacciones algorítmicas en contextos económicos dinámicos y cambiantes.

2.1.8.2. Modelamiento

Inspirado en la metodología de trabajo expuesta en [11], se define el nivel de Precios general de la economía λ_t del cual se deriva la expresión:

$$\omega_t = \frac{\lambda_t - \lambda_{t-1}}{\lambda_{t-1}} \tag{2.11}$$

donde ω_t representa el crecimiento del nivel de precios en la economía o *inflación* de precios.

Considerando la expresión anterior y asumiendo que los costos de producción dependen directamente del nivel de Precios λ_t, es posible modelar el aumento en los costos de producción mediante la siguiente identidad:

$$c_t = c_{t-1} \cdot (1 + \omega_t) \tag{2.12}$$

donde π_t representa el shock inflacionario determinado de forma exógena a las acciones de las firmas.

Para determinar el valor de π_t y con tal de preservar una cuota de realismo en los experimentos, se utiliza una muestra de diferentes series inflacionarias mensuales de un pool de 10 países entre 2000 y 2015, los cuales fueron seleccionados en función de su estabilidad en el crecimiento de su nivel de precios y de esta manera ayudar en la estabilización de los gradientes en los experimentos. El Anexo A.1 presenta una caracterización estadística de la distribución de las series utilizadas.

En consideración de lo expuesto en la sección, es necesario recalibrar la función de demanda de la ecuación 2.7 por medio del modelamiento del crecimiento en la disposición a pagar de los consumidores. Siguiendo el trabajo de [12], la disposición a pagar de los consumidores $\alpha_{i,t}$ depende de dos grandes factores: el ingreso percibido y_t y las preferencias individuales de los consumidores u_t.

$$\alpha_{i,t}(y_t, u_t) \tag{2.13}$$

Para los propósitos de este estudio, se postula que las preferencias de los consumidores se mantienen invariables a lo largo del tiempo, lo que implica que estos asignarán una fracción constante de su ingreso disponible a la adquisición de dichos productos a lo largo de las distintas etapas temporales. Por otra parte, con el objetivo de preservar la simplicidad del análisis, se presupone que el ingreso disponible aumenta a la misma velocidad que costo de producción de las empresas. Este último supuesto podría considerarse poco realista, especialmente dado que el crecimiento de la inflación y los salarios no ha seguido una trayectoria homogénea en los últimos años (ver Figura 2.4). Explorar de manera independiente el crecimiento de los precios y la disposición a pagar emerge como una posible vía para ampliar la investigación en este ámbito.

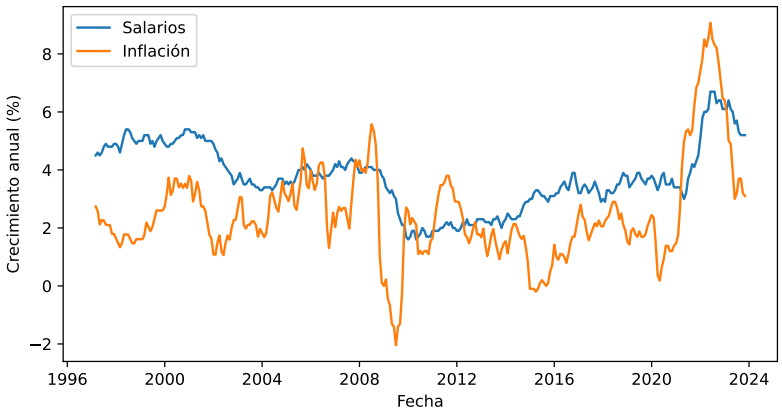


Figura 2.4: Diferencia entre la tasa de inflación y el crecimiento de los salarios en Estados Unidos desde Marzo 1997 hasta Noviembre 2023.

Dado lo anterior y por simplicidad en el estudio, se asume que el crecimiento del factor $\alpha_{i,t}$ sea determinado por la siguiente ecuación:

$$\alpha_{i,t} = \alpha_{i,t-1} \cdot (1 + \omega_t) \tag{2.14}$$

donde $\alpha_{i,t}$ representa la disposición a pagar de los consumidores hacia el producto de la firma i en el período t y π_t el nivel de inflación en el período t.

2.2. Aprendizaje Reforzado

2.2.1. Introducción

El aprendizaje reforzado emerge como un paradigma en el ámbito de la inteligencia artificial, donde un agente, al enfrentarse a un entorno dinámico, internaliza la toma de decisiones secuenciales con la finalidad de maximizar las recompensas acumulativas. En contraste con el aprendizaje supervisado que se fundamentan en conjuntos de datos etiquetados y los métodos no supervisados centrados en la identificación de patrones sin guía explícita, el aprendizaje reforzado se caracteriza por la adquisición autónoma de estrategias óptimas mediante la interacción directa con el entorno y la retroalimentación derivada de recompensas o penalizaciones. Esta habilidad innata de aprendizaje adaptativo otorga a esta disciplina una utilidad trascendental en la resolución de problemáticas complejas y dinámicas. Más allá de sus aplicaciones fundamentales en sectores como la robótica y el control de procesos, el aprendizaje reforzado destaca en la modelización de comportamientos estratégicos en el marco de la teoría de juegos, ofreciendo una perspectiva esencial para comprender estrategias óptimas en interacciones estratégicas. Su idoneidad para abordar problemas de toma de decisiones secuenciales y adaptarse a entornos en constante cambio le confiere un estatus preeminente en el ámbito de la inteligencia artificial, con repercusiones significativas tanto en contextos aplicados como en investigaciones teóricas.

Siguiendo lo expuesto en [13], todo problema de Aprendizaje Reforzado puede ser representado a través de un **Proceso de Decisión de Markov**, el cual a su vez se compone de:

- **Agente:** El agente representa el sistema de toma de decisiones en el contexto del Aprendizaje Reforzado. Es el componente que interactúa con el entorno, toma acciones y busca desarrollar políticas óptimas para maximizar las recompensas acumulativas a lo largo del tiempo.
- **Acciones:** Las acciones representan las decisiones específicas que el agente puede tomar en un estado determinado del entorno. El conjunto de acciones disponibles influye directamente en la capacidad del agente para explorar y explotar el entorno de manera efectiva. La selección adecuada de acciones es fundamental para la optimización de las políticas y la maximización de las recompensas a lo largo del tiempo. Las acciones pueden ser discretas o continuas, dependiendo de la naturaleza del problema.
- **Ambiente:** El entorno constituye el contexto en el que opera el agente. Es el sistema externo que responde a las acciones del agente y proporciona retroalimentación en forma de estados y recompensas.
- **Estados:** Los estados representan las distintas situaciones o configuraciones en las que puede encontrarse el entorno. Son observaciones del entorno que el agente utiliza para tomar decisiones. La dinámica entre estados, acciones y recompensas es fundamental para el desarrollo de estrategias de aprendizaje efectivas en el contexto del Aprendizaje Reforzado.
- **Recompensas:** Las recompensas cuantifican la consecuencia positiva o negativa de las acciones tomadas por el agente en un estado particular. Son esenciales para guiar al

agente hacia la toma de decisiones que maximice su recompensa acumulativa a lo largo de las interacciones con el entorno.

- **Políticas:** Las políticas son estrategias que guían al agente en la selección de acciones en diferentes estados del entorno. El objetivo es desarrollar políticas óptimas que lleven al agente a tomar decisiones que maximicen las recompensas a largo plazo. La exploración y explotación son aspectos clave en el desarrollo de políticas efectivas. En términos matemáticos, puede ser interpretada como una función que mapea desde estados a la probabilidad de tomar cada acción disponible.

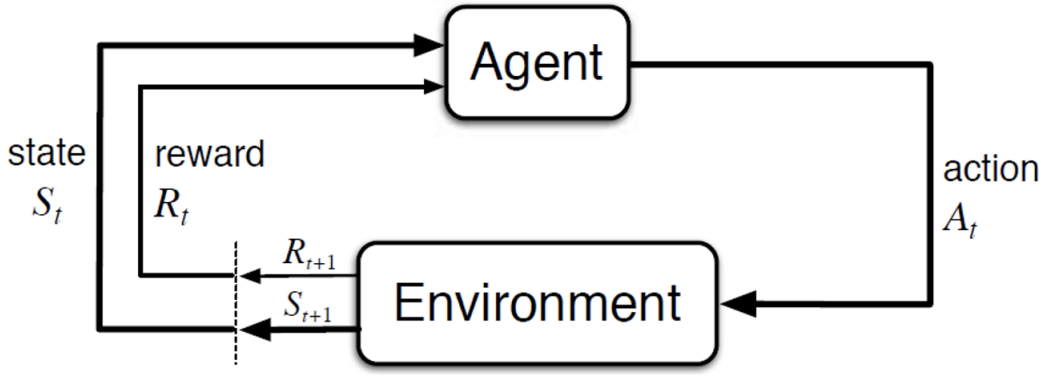


Figura 2.5: Elementos fundamentales de Aprendizaje Reforzado (figura extraida de [14]).

Finalmente, un problema de Aprendizaje Reforzado se considera resuelto cuando se encuentra aquella política que maximiza la recompensa del agente a lo largo del experimento. En otras palabras, el problema de optimización a resolver puede ser expresado por:

$$\max_{\pi} J(\pi) = \mathbb{E}_{\tau \sim p_\pi(\tau)} \left[\sum_{t=1}^{T} \gamma^{t-1} r\left(s_t, a_t \sim \pi(a|s)\right) \right] \tag{2.15}$$

donde π corresponde a la política del agente, τ expresa la dinámica de transición entre estados, r indica la recompensa del agente dado el estado s_t y su accion a_t, y γ indica la tasa de descuento intertemporal.

2.2.2. Métodos para resolver problemas de RL

De acuerdo a [13], los métodos de aprendizaje por refuerzo se pueden categorizar en dos enfoques fundamentales: métodos basados en el valor (o *Value-Based*) y métodos basados en la política (*Policy-Based*). Estas categorías se diferencian principalmente en cómo abordan la toma de decisiones del agente y la actualización de su estrategia. A continuación, se entrega una descripción detallada de ambos enfoques:

- **Value-Based**: En los métodos basados en el valor, el objetivo es aprender y representar la función de valor de un estado y acción, denotada comúnmente como $Q(s, a)$. La función de valor refleja la utilidad esperada asociada de tomar una acción a en un estado s en términos de la recompensa acumulada. La ecuación de Bellman para $Q(s, a)$ en términos

de la recompensa inmediata y la función de valor usando el estado siguiente s' es una expresión clave en estos métodos:

$$Q(s,a) = r(s,a) + \gamma \max_a Q(s',a) \tag{2.16}$$

Donde s y a representan el estado actual y la acción tomada, respectivamente, s' es el siguiente estado, γ es el factor de descuento que pondera las recompensas futuras, y $\max_a Q(s',a)$ denota la estimación del valor máximo del próximo estado. Algunos algoritmos notables de este enfoque son Q-learning [15] y Deep Q-Network [16].

- **Policy-Based**: En contraste, los métodos basados en la política buscan aprender directamente la política de decisión del agente, es decir, la distribución de probabilidad sobre las acciones condicionadas al estado. La política es denotada como $\pi(a|s)$, que representa la probabilidad de elegir la acción a dado el estado s. La actualización de la política se realiza mediante gradientes, y la ecuación de la política es fundamental en este enfoque:

$$\nabla_\theta J(\theta) \propto \sum_s \mu(s) \sum_a \nabla_\theta \pi(a|s,\theta) Q^\pi(s,a) \tag{2.17}$$

Donde $J(\theta)$ representa el rendimiento esperado de la política, $\mu(s)$ es la distribución estacionaria de estados bajo la política actual, $\pi(a|s,\theta)$ es la política parametrizada por θ y $Q^\pi(s,a)$ es la función de valor bajo la política π. Ejemplos notables de este enfoque son los algoritmos REINFORCE [17] y TRPO [18].

2.2.3. Configuración Multiagente

El Aprendizaje Reforzado Multiagente (MARL) representa una extensión integral del paradigma de Aprendizaje Reforzado, diseñada específicamente para abordar la complejidad intrínseca de entornos donde múltiples agentes interactúan simultáneamente. A diferencia del Aprendizaje Reforzado clásico, donde un solo agente busca optimizar sus acciones en un entorno dado, el MARL considera la existencia de múltiples agentes, cada uno con sus propias metas y objetivos individuales. Esta interacción introduce una dinámica estratégica más compleja, ya que las decisiones de un agente impactan directamente en el entorno y, por ende, en las oportunidades y resultados disponibles para los demás agentes.

Los principales componentes del aprendizaje reforzado multiagente incluyen a los agentes, el ambiente compartido, las acciones, los estados y las recompensas asociadas con las interacciones. La Figura 2.6 ilustra estos elementos en un contexto de MARL [19].

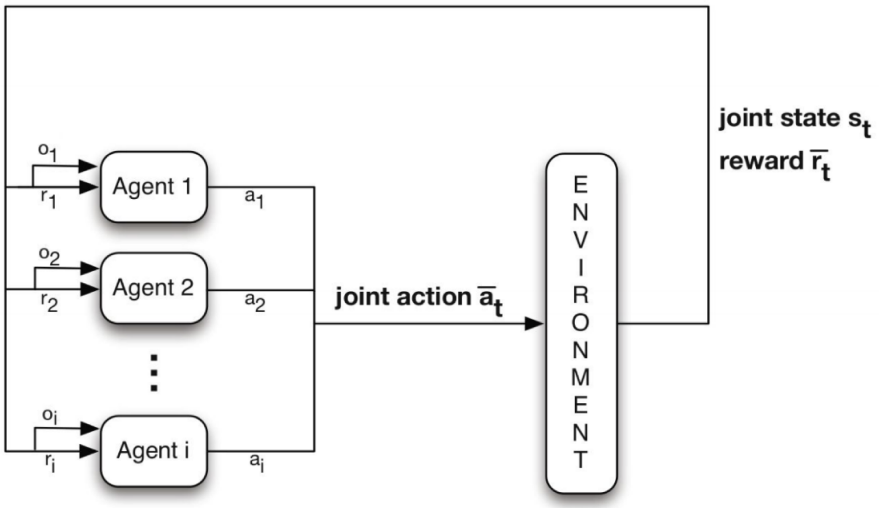


Figura 2.6: Elementos de Aprendizaje Reforzado Multiagente.

La interdependencia estratégica entre agentes en el Aprendizaje Reforzado Multiagente (MARL) introduce desafíos adicionales relacionados con la coordinación, la competencia y la formación de coaliciones, elevando la complejidad del problema. A diferencia del Aprendizaje Reforzado convencional, donde la toma de decisiones se realiza en un entorno relativamente aislado, el MARL exige que cada agente considere las acciones y estrategias de los demás para alcanzar sus propios objetivos. Esta necesidad de coordinación y adaptación constante entre agentes refleja una interdependencia estratégica que define la dinámica única del MARL.

La interdependencia entre agentes no solo desafía la propiedad de Markov al introducir interacciones complejas y dinámicas en un entorno compartido, sino que también resalta la naturaleza no estacionaria del aprendizaje reforzado multiagente. En particular, la propiedad de Markov, donde las recompensas, dinámica y transiciones dependen exclusivamente del estado actual, deja de ser válida en este contexto, ya que ahora estas variables también están condicionadas por las acciones de los demás agentes [20]. Esta pérdida de estacionariedad conlleva a la falta de garantía de convergencia hacia una política óptima, marcando un desafío sustancial en la aplicación del aprendizaje reforzado multiagente en entornos complejos e interactivos. A pesar de la falta de garantías, los aprendices independientes se han utilizado en la práctica, ofreciendo ventajas en cuanto a escalabilidad y, en muchas ocasiones, logrando buenos resultados [21].

Por último, el Aprendizaje Reforzado Multiagente encuentra aplicaciones prácticas en diversos campos, como juegos estratégicos, sistemas multirobot y redes de comunicación. Su relevancia en teoría de juegos se manifiesta en la capacidad de modelar y analizar interacciones estratégicas complejas entre agentes racionales. El MARL permite explorar estrategias y dinámicas emergentes, ofreciendo un marco analítico poderoso para entender cómo las decisiones de los agentes individuales convergen hacia resultados colectivos en contextos competitivos o colaborativos. En este sentido, el Aprendizaje Reforzado Multiagente se presenta como una disciplina esencial para abordar la complejidad de sistemas interactivos, destacando su potencial para informar y enriquecer la teoría de juegos.

2.2.4. Aprendizaje Reforzado Profundo

El Aprendizaje Reforzado Profundo [22] representa una evolución significativa con respecto a los métodos tradicionales de Aprendizaje Reforzado, tales como el aprendizaje basado en Q-learning, especialmente en dominios donde el problema de la maldición de la dimensionalidad se torna especialmente relevante. Los métodos tabulares, aunque eficaces en entornos con espacios de estados restringidos, enfrentan desafíos notables en términos de velocidad de aprendizaje, una generalización deficiente a lo largo del espacio de estados y la necesidad de definir manualmente las representaciones de estado.

Esta avanzada modalidad de aprendizaje aborda estas limitaciones mediante la incorporación de aproximadores de funciones más flexibles, destacando entre ellos las redes neuronales. La representación de los valores de estado-acción mediante una red neuronal $Q(s, a; \theta)$ ofrece dos beneficios fundamentales. Primero, el aprendizaje profundo facilita una mejor generalización a través de distintos estados, mejorando así la eficiencia del muestreo en problemas de Aprendizaje Reforzado con amplios espacios de estado. Segundo, reduce o elimina la necesidad de diseñar manualmente características para describir la información del estado, simplificando el proceso de modelado.

No obstante, la incorporación del aprendizaje profundo en los problemas de Aprendizaje Reforzado introduce retos adicionales, como la naturaleza no independiente e idénticamente distribuida de los datos. A diferencia de los métodos convencionales de aprendizaje supervisado, que presuponen una distribución estacionaria y uniforme, en el Aprendizaje Reforzado, los datos de entrenamiento consisten en interacciones secuenciales y altamente correlacionadas entre el agente y su entorno, lo cual viola esta suposición. Además, la distribución de los datos de entrenamiento en el Aprendizaje Reforzado es dinámica, debido a que el agente aprende activamente mientras explora diferentes partes del espacio de estados, alterando así la premisa de que los datos muestreados sean uniformes.

En lo que respecta a técnicas específicas, el Aprendizaje Reforzado Profundo ha introducido estrategias innovadoras como el *Replay Buffer* [23] y las *Target Network* [24]. El *Replay Buffer* afronta la correlación secuencial en los datos de entrenamiento, permitiendo al agente almacenar y reutilizar interacciones pasadas. Por su parte, las *Target Network* atenúan la inestabilidad en la actualización de la red neuronal objetivo al introducir una red de seguimiento más lenta que se adhiere a parámetros más estables, contribuyendo así a la estabilidad del proceso de entrenamiento. Estas técnicas han sido fundamentales para mejorar la eficacia y la estabilidad del Aprendizaje Reforzado Profundo en problemas complejos de Aprendizaje Reforzado.

2.3. Revisión de Literatura

En el campo de la fijación de precios algorítmica, varios estudios recientes han buscado contribuir a la literatura de *Algorithmic Collusion*. En el estudio de Calvano et al. (2020) [25] demostraron que los algoritmos de aprendizaje por refuerzo pueden aprender estrategias colusivas en mercados simulados, donde los precios son actualizados simultáneamente por competidores. Este trabajo sugiere que la colusión puede surgir incluso sin una comunicación explícita entre los agentes, siempre que los algoritmos puedan observar los precios pasados de sus competidores y ajustarse en consecuencia.

Lepore (2021) [26] investigó la colusión de precios utilizando algoritmos de aprendizaje por refuerzo en un entorno de competencia de Bertrand. Este estudio replicó y extendió los resultados de Calvano et al., mostrando que algoritmos más complejos pueden coludir de manera más confiable y rápida en comparación con algoritmos simples. Además, se exploraron métodos para mitigar esta colusión mediante la introducción de un agente supervisor que influye en la demanda, similar a la técnica de la "carro de compra" de Amazon.

Klein (2021) [3] analizó la colusión algorítmica en un entorno de precios secuenciales, utilizando Q-learning. Su investigación demostró que los algoritmos competidores pueden coordinarse en equilibrios colusivos, especialmente cuando el número de precios discretos es limitado. Este estudio destaca cómo los algoritmos aprenden a coordinarse para mantener precios altos, lo que plantea desafíos significativos para la regulación y el diseño de políticas que buscan evitar la colusión.

Eschebaum et al. (2021) [27] proporcionaron una perspectiva crítica sobre la evaluación de la colusión algorítmica, subrayando la importancia de un entorno de prueba separado del entorno de entrenamiento para evaluar adecuadamente el comportamiento de los algoritmos. Este trabajo sugiere que la tendencia de los algoritmos a sobreajustarse al entorno de entrenamiento puede influir significativamente en su capacidad para coludir en la práctica, destacando la necesidad de enfoques más robustos para la evaluación de algoritmos en juegos económicos.

Finalmente, Abada y Lambin (2022) [28] analizaron la colusión algorítmica en el contexto de sistemas de energía descentralizados. Sus hallazgos indican que la colusión puede surgir más fácilmente de lo sugerido en estudios anteriores, y que las intervenciones regulatorias durante el proceso de aprendizaje pueden mitigar los efectos destructivos para el bienestar de la colusión aparente.

Capítulo 3

Metodología

3.1. Proceso de Decisión de Markov

Esta sección contiene la formulación de lo expuesto en la sección 2.1 como un Proceso de Decisión de Markov.

3.1.1. Ambiente

El crecimiento de los costos es modelado a través de un factor inflacionario π_t:

$$c_t = c_{t-1} \cdot (1 + \omega_t) \tag{3.1}$$

Donde la tasa de inflación ω_t representa la tasa de cambio en los costos de producción c_t a través del tiempo. Esta tasa es determinada de forma estocástica y exógena extrayendo cifras de series inflacionarias mensuales de diferentes países:

$$\omega_t = \begin{cases} \bar{\omega}_i & \text{con probabilidad } \rho \\ 0 & \text{en otro caso} \end{cases} \tag{3.2}$$

donde $\bar{\pi}_i$ representa la i-ésima observación de la serie inflacionaria mensual escogida de forma aleatoria y ρ la probabilidad de ocurrencia de el aumento en los precios. El modelamiento estocástico del shock inflacionario es útil por 2 razones: (i) permite que los resultados de los experimentos puedan ser extrapolados a una unidad de tiempo interpretable por medio del ajuste del parámetro ρ (donde el caso mas simple ocurre cuando $\rho = 1$ y cada *timestep* puede ser interpretado como un mes); (ii) otorga un mayor realismo a los experimentos pues en la práctica las empresas no conocen el *timing* del shock en el aumento de los costos.

Con las ecuaciones 3.1 y 3.2, es posible definir un índice de precios λ_t que cuantifique el nivel de precios en el tiempo t:

$$\lambda_t = \lambda_{t-1} \cdot (1 + \omega_t) \tag{3.3}$$

Considerando la ecuación 3.3 y el modelo económico establecido en el capítulo 2, se define la dinámica del ambiente a través de la función de demanda siguiendo el modelo de Bertrand con productos diferenciados e inflación de precios a partir de la siguiente expresión:

$$q_{i,t} = \frac{e^{\frac{\alpha_i - p_{i,t}}{\lambda_t \cdot \mu}}}{\sum_{j=1}^{n} e^{\frac{\alpha_j - p_{j,t}}{\lambda_t \cdot \mu_i}} + e^{\frac{\alpha_0}{\lambda_t \cdot \mu}}} \tag{3.4}$$

Finalmente, con tal de asegurar la prevalencia de los mercados y según lo discutido en la sección 2.1.8, se impone el crecimiento de la disposición a pagar en una tasa igual a la tasa de inflación, es decir:

$$\alpha_{i,t} = \alpha_{i,t-1} \cdot (1 + \omega_t) \tag{3.5}$$

3.1.2. Recompensas

Siguiendo la formulación de cualquier problema de maximización de beneficios en la teoría económica, el beneficio económico de una empresa puede ser formulado como la resta entre los ingresos y los costos como consecuencia del precio fijado $p_{i,t}$, es decir:

$$R_{i,t} = (p_{i,t} - c_t) \cdot q_{i,t} \tag{3.6}$$

Como la estacionariedad en las utilidades es algo deseable para la convergencia de los algoritmos, se *deflactan* las utilidades por el nivel de precios λ_t. Por lo tanto, las recompensas de cada agente vienen dadas por:

$$R_{i,t} = \frac{(p_{i,t} - c_t) \cdot q_{i,t}}{\lambda_t} \tag{3.7}$$

3.1.3. Acciones

Recordando los supuestos del modelo de Bertrand expuestos en la sección 2.1, los agentes interactuán con la demanda mediante la fijación del precio de venta de sus productos. En otras palabras, es posible interpretar el precio a fijar como la acción emitida por el agente en el Proceso de Decisión de Markov.

Para cada período t, cada agente es capaz de emitir un precio $p_{i,t}$, el cual tiene la restricción de pertenecer a los reales positivos:

$$p_{i,t} \in \mathbb{R}^+ \tag{3.8}$$

Siguiendo la metodología de [2], el espacio de acciones posibles es discreto y se limita a m posibles precios, los cuales son calculados de manera equidistante entre cotas predefinidas, tomando usualmente como referencia el precio de Nash p^N y el precio de Monopolio p^M:

$$p_{i,t} \in \left[p^N - \xi\left(p^M - p^N\right), p^M + \xi\left(p^M - p^N\right)\right] \tag{3.9}$$

donde $0 \leq \xi \leq 1$. Como estos precios son desconocidos para las empresas en un escenario real, se redefine el espacio de acciones a *precios sobre costos*, es decir:

$$p_{i,t} = c_t \cdot (1 + \eta_{i,t}) \tag{3.10}$$

Donde $\eta_{i,t}$ es una de las m opciones equidistantes entre η_{min} y η_{max} escogida por el agente i para el período t. En términos prácticos, $\eta_{i,t}$ es determinado por:

$$\eta_{i,t} = \eta_{min} + a_{i,t} \cdot \frac{\eta_{max} - \eta_{min}}{m - 1} \tag{3.11}$$

donde $a_{i,t} \in (0, m-1)$ representa la salida de la políca del agente i en el *timestep* t.

Bajo esta formulación, $\eta_{i,t}$ puede ser interpretado como el *margen sobre costos objetivo* al cual apunta el agente i para el período t:

$$\eta_{i,t} = \frac{p_{i,t} - c_t}{c_t} \tag{3.12}$$

3.1.4. Estados

Siguiendo la metodología de [2] y en función de los objetivos propuestos en la sección 1.3, es posible otorgar *memoria* a los agentes para fijar su precio. Con esto en mente, es posible modelar las observaciones de los agentes como el precio fijado por los agentes en los últimos k períodos:

$$S_t = \{P_{t-k}, ..., P_{t-1}\} \tag{3.13}$$

donde P_t indica el conjunto de precios $\{p_{i,t}, ..., p_{j,t}\}$ fijados por todos los agentes en el timestep t.

Considerando que la magnitud de los precios aumenta progresivamente como consecuencia del aumento en los costos, se reformulan los estados *deflactando* los precios por el nivel de costos, es decir:

$$S_t = \left\{ \frac{P_{t-k} - c_{t-k}}{c_{t-k}}, ..., \frac{P_{t-1} - c_{t-1}}{c_{t-1}} \right\} \tag{3.14}$$

Esta formulación es conveniente pues permite limitar el crecimiento en la magnitud de los estados a un intervalo acotado, permitiendo una mejor estabilización de los gradientes de los agentes. En particular, la vecindad de esta expresión queda acotada a los umbrales porcentuales η_{min} y η_{max} de acuerdo a la ecuación 3.12. De esta manera, los agentes observan el *margen* sobre los costos fijado en los k períodos anteriores.

Finalmente y con tal de que los agentes puedan reconocer la ocurrencia de un shock inflacionario en el timestep t, es conveniente añadir el valor de los costos c_t para el período t y sus valores en los k períodos anteriores. De esta forma, el estado recibido por los agentes puede ser expresado por:

$$S_t = \{c_t\} \cup \left\{ \frac{P_{t-k} - c_{t-k}}{c_{t-k}}, ..., \frac{P_{t-1} - c_{t-1}}{c_{t-1}} \right\} \cup \{c_{t-k}, ..., c_{t-1}\} \tag{3.15}$$

donde P_t representa el conjunto de precios fijados por las n empresas en competencia en el periodo t, c_t hace alusión al costo de producción unitario de los agentes en el periodo t y k representa el número de periodos pasados a los que los agentes tienen acceso.

3.1.5. Diagrama de Ambiente

Con tal de ilustrar el ambiente descrito a lo largo de esta sección, la Figura 3.1 representa el flujo de entrenamiento para 2 agentes.

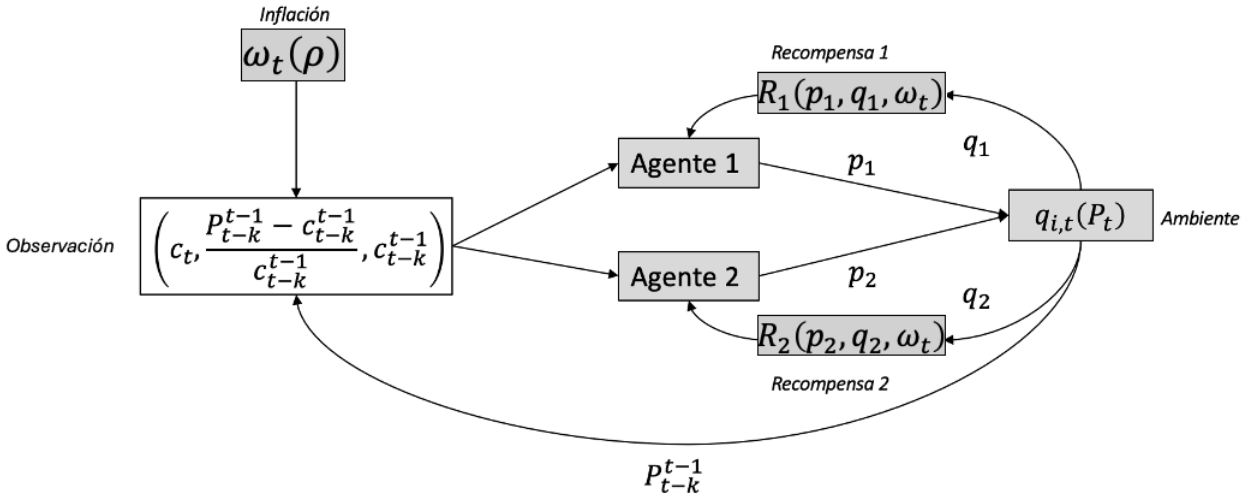


Figura 3.1: Flujo de entrenamiento para dos agentes.

3.2. Experimento con Agente Altruista

El altruismo, como constructo en la teoría económica, se refiere a la preocupación por el bienestar de los demás, a menudo en detrimento del interés personal. A diferencia del modelo tradicional de *homo economicus*, que postula un comportamiento estrictamente autocentrado y maximizador de utilidades, el agente altruista incorpora el bienestar de los demás en su función de utilidad [29]. Este enfoque desafía la noción de racionalidad egoísta y sugiere una dimensión más compleja y matizada de la toma de decisiones humanas [30].

En un contexto de oligopolio, donde la colusión entre firmas es común para maximizar beneficios a expensas del consumidor, la introducción de un agente altruista puede alterar significativamente la dinámica del mercado. Al priorizar el bienestar colectivo, un agente altruista puede optar por estrategias que promuevan una mayor competencia, subviertan acuerdos colusorios y fomenten la reducción de precios, beneficiando así a los consumidores y aumentando la eficiencia del mercado [31]. Esta presencia desafía la estabilidad de la colusión y puede inducir a otros actores del mercado a adaptar estrategias menos centradas en el monopolio y más orientadas hacia el bienestar social.

Considerando lo anterior y con el espíritu de buscar posibles alternativas para mitigar el impacto negativo del pricing algorítimico sobre la competitividad de los mercados, se propone la generación de un experimento con la participación de un agente altruista, similar a una empresa estatal que persigue el beneficio de la sociedad. Este experimento comparte

las misma formulación que las detalladas en la sección 3.1, con la única diferencia de que el agente altruista siempre fijará su precio igual al precio de Nash, es decir:

$$p_{0,t} = p_{N,t} \; \forall \; t \tag{3.16}$$

donde en este caso el agente 0 asume el rol de "altruista" y $p_{N,t}$ representa el precio de *Nash* calculado para el péríodo t.

3.2.1. Diagrama de Ambiente

Con tal de ilustrar el experimento con la participación de un agente altruista, la Figura 3.2 ilustra el flujo de entrenamiento bajo la configuración expuesta:

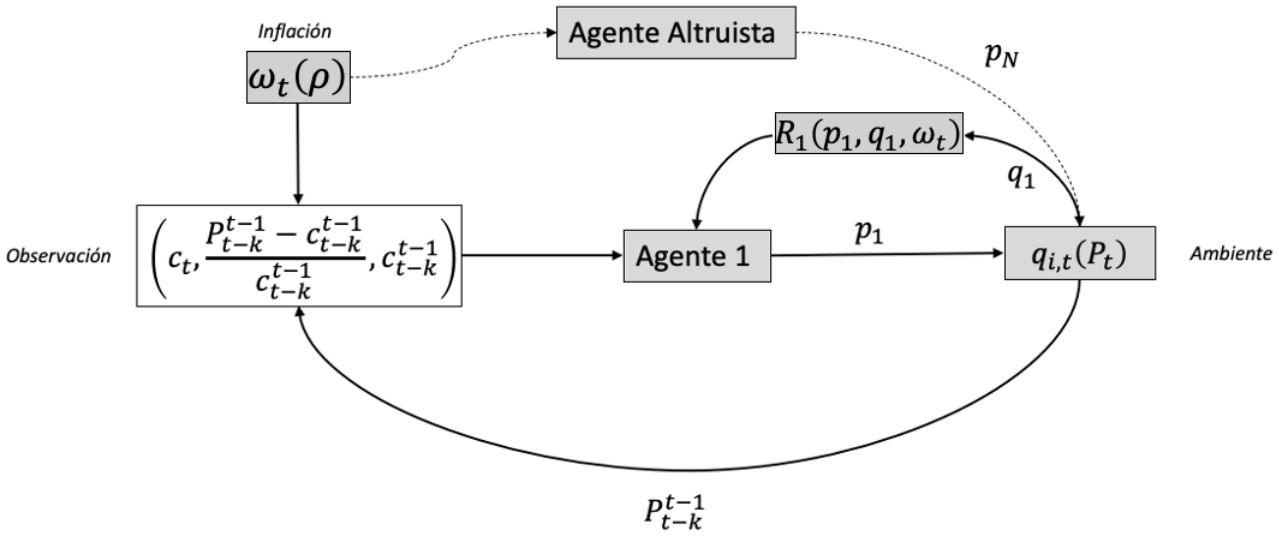


Figura 3.2: Ciclo de entrenamiento con agente altruista.

3.3. Agente

En esta sección se presenta el mecanismo por el que los Agentes obtienen una política para emitir acciones en el ambiente señalado.

3.3.1. Deep Q-Network

El algoritmo **Deep Q-Network** (DQN) es una innovación clave en el campo del Aprendizaje Reforzado, introducido por Volodymyr Mnih y otros en el estudio titulado "Playing Atari with Deep Reinforcement Learning" (2013) [32]. Su formulación matemática se basa en la representación de la función de valor óptima, denotada por $Q^*(s, a)$ la cual predice la recompensa acumulada futura al tomar la acción a en el estado s.

La ecuación de Bellman, fundamental en RL, guía la actualización de la función $Q*$:

$$Q^*(s,a) = \mathbb{E}\left[R_t + \gamma \max_{a'} Q^*(s',a')|s,a\right] \tag{3.17}$$

Donde R_t es la recompensa en el tiempo, s' es el siguiente estado, γ es el factor de descuento, y la esperanza $\mathbb{E}$ se toma sobre las posibles transiciones del entorno.

DQN utiliza una red neuronal profunda para aproximar la función Q^*. Durante el entrenamiento, se minimiza el error cuadrático medio entre la predicción de la red y el objetivo de la ecuación de Bellman. Para estabilizar el aprendizaje, se emplean técnicas como la experiencia por repetición (*replay buffer*) y una red objetivo (*target network*), abordando desafíos asociados con la correlación temporal y la no estacionariedad de los datos de entrenamiento.

Las ventajas de DQN incluyen la capacidad para manejar espacios de estado continuos y de alta dimensionalidad, así como la generalización efectiva a partir de grandes conjuntos de datos. Sin embargo, sufre de sobreestimación de valores (pueden surgir estimaciones exageradas), y la elección de hiperparámetros puede ser crucial para el rendimiento óptimo.

Comparado con métodos tradicionales, DQN ha demostrado un rendimiento notable en tareas complejas y ha allanado el camino para enfoques más avanzados en el ámbito del Aprendizaje Reforzado Profundo. La Figura 3.3 presenta el pseudocódigo del algoritmo DQN, donde se utiliza la estrategia *Epsilon-greedy* para tratar el dilema de exploración-explotación de estados.

Initialize replay memory $\mathcal{D}$ to capacity N
Initialize action-value function Q with random weights
for episode $= 1, M$ **do**
  Initialise sequence $s_1 = \{x_1\}$ and preprocessed sequenced $\phi_1 = \phi(s_1)$
  for $t = 1, T$ **do**
    With probability ϵ select a random action a_t
    otherwise select $a_t = \max_a Q^*(\phi(s_t), a; \theta)$
    Execute action a_t in emulator and observe reward r_t and image x_{t+1}
    Set $s_{t+1} = s_t, a_t, x_{t+1}$ and preprocess $\phi_{t+1} = \phi(s_{t+1})$
    Store transition $(\phi_t, a_t, r_t, \phi_{t+1})$ in $\mathcal{D}$
    Sample random minibatch of transitions $(\phi_j, a_j, r_j, \phi_{j+1})$ from $\mathcal{D}$
    Set $y_j = \begin{cases} r_j & \text{for terminal } \phi_{j+1} \\ r_j + \gamma \max_{a'} Q(\phi_{j+1}, a'; \theta) & \text{for non-terminal } \phi_{j+1} \end{cases}$
    Perform a gradient descent step on $(y_j - Q(\phi_j, a_j; \theta))^2$
  end for
end for

Figura 3.3: Pseudocódigo de algoritmo DQN presentado en [32].

3.3.2. Epsilon-greedy

En el campo del aprendizaje reforzado, la estrategia *Epsilon-greedy* se erige como un enfoque fundamental. Esta estrategia juega un papel crucial para equilibrar la exploración de nuevas acciones frente a la explotación de las ya conocidas, facilitando que un agente aprenda a actuar de manera óptima en un entorno desconocido. La estrategia *Epsilon-greedy* es esencial para determinar cuándo un agente debe explorar nuevas posibilidades en vez de aprovechar el conocimiento previamente adquirido.

La formulación matemática de la estrategia *Epsilon-greedy* se define de la siguiente manera:

$$a_t = \begin{cases} \text{acción aleatoria}, & \text{con probabilidad } \epsilon_t \\ \underset{a}{\operatorname{argmax}}\, Q(s_t, a), & \text{con probabilidad } 1 - \epsilon_t \end{cases} \tag{3.18}$$

Donde ϵ_t es un parámetro que determina la frecuencia de exploración, a_t es una de las m acciones posibles escogida en el tiempo t y $Q(s_t, a)$ es la función que estima la recompensa de una acción a en un estado s_t. Siguiendo la metodología de [2], se adopta una reducción progresiva del factor ϵ_t, permitiendo una fase inicial de exploración intensiva seguida de un enfoque más concentrado en la maximización de la recompensa. La dinámica de decaimiento de ϵ se modela con la ecuación:

$$\epsilon_t = e^{-\beta t} \tag{3.19}$$

donde β controla la velocidad a la que el parámetro ϵ_t disminuye.

En términos de ventajas, la estrategia *Epsilon-greedy* se destaca por su simplicidad y flexibilidad, adaptándose a una variedad de contextos y problemas y proporcionando un balance efectivo entre exploración y explotación. Sin embargo, esta estrategia no está exenta de desventajas. A pesar de la adaptabilidad del factor ϵ_t a través del parámetro β, la elección aleatoria de acciones puede aún conducir a decisiones subóptimas, especialmente en espacios de acción amplios donde la probabilidad de seleccionar la acción óptima aleatoriamente es baja. Además, la estrategia no considera la calidad o el potencial de las acciones no exploradas, lo que podría conducir a un aprendizaje ineficiente en ciertos contextos.

3.3.3. Arquitectura Neuronal

En el contexto del estudio sobre *Algorithmic Collusion* utilizando agentes de aprendizaje reforzado, resulta crucial la implementación de una red neuronal, denotada como ϕ_θ. Conforme a lo expuesto en las sección 3.3.1, esta red está diseñada para procesar el estado s_t y aproximar de manera eficiente la función de valor $Q(s_t, a)$. Esta relación se formaliza mediante la siguiente expresión matemática:

$$\phi_\theta(s_t) \rightarrow Q(s_t, a) \tag{3.20}$$

Para este propósito, se configura cada agente como un Perceptrón Multicapa (MLP) compuesto por tres capas ocultas. Entre cada una de estas capas se incorpora la función de activación ReLU, elegida por su eficacia en prevenir el problema del desvanecimiento del gradiente, aspecto crucial en el aprendizaje profundo. Antes de su introducción en la red neuronal, se realiza una operación *flatten* sobre los componentes del estado s_t, transformándolos en un formato adecuado para el procesamiento por el MLP.

Posteriormente, los componentes aplanados se concatenan formando un vector de dimensiones $N \times k + k + 1$. Este vector se somete a un proceso de normalización utilizando una media móvil con un tamaño de $1/\rho$, un paso vital para mitigar los problemas asociados con la estacionariedad en series temporales de datos. La Figura 3.4 proporciona una representación gráfica detallada de la arquitectura neuronal implementada:

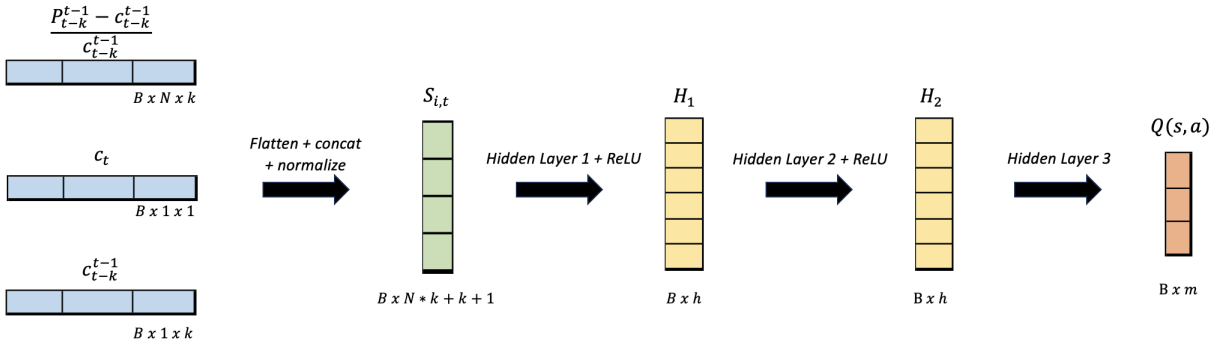


Figura 3.4: Arquitectura Neuronal de Agentes.

donde B indica el tamaño de cada *batch* muestreado desde el *Replay Buffer*, h indica el número de neuronas de las capas ocultas, m indica el número de acciones disponibles y H_i indica el estado oculto generado tras transformar datos con la capa lineal i.

3.4. Diseño Experimental

3.4.1. Experimentos

Para validar la hipótesis de investigación, se diseñan los siguientes experimentos:

- **Experimento base**: Este experimento tiene como objetivo evaluar el impacto de incluir variabilidad en los costos de producción en los agentes sobre los fenómenos identificados en la literatura de *Algorithmic Collusion*. Para lograr lo anterior, se utilizan agentes basados en el algoritmo DQN bajo un conjunto de hiperparámetros base, los cuales fueron elegidos en función de los estudios [2], [26] y [16]. La configuración de estos parámetros se resume en la Tabla 3.1:

Tabla 3.1: Configuración base.

Hiperparámetro	**Valor**
Número de Agentes (N)	2
Cantidad de períodos pasados (k)	1
Probabilidad de shock inflacionario (ρ)	0.001
Tasa de aprendizaje (lr)	0.01
Factor de descuento (γ)	0.95
Mínima Variación sobre costos (η_{min})	-0.5
Máxima Variación sobre costos (η_{max})	2.0
Número de acciones (m)	15
Optimizador	Adam [33]
Costo de producción inicial ($c_{t=0}$)	1
Diferenciación vertical inicial ($\alpha_{t=0}$)	1
Índice inverso de demanda agregada (α_0)	0
Diferenciación horizontal (μ)	0.25
Número de neuronas por capa (h)	256
Capas ocultas	2
Tamaño de batch (B)	256
Tamaño Replay Buffer	20.000
Episodios	1
Timesteps	400.000
Pasos para calcular gradiente	1
Pasos para actualizar Target	200

Es importante resaltar el hecho de que la negatividad del valor de η_{min} permite que los agentes puedan fijar precios bajo sus costos de producción c_t, otorgando una dificultad adicional en el experimento para aprender a fijar un precio mayor o igual a estos costos.

A modo de contraste, se buscará repetir el experimento con la misma configuración, pero sin la inclusión de variabilidad en los costos (es decir, $\gamma = 0$). De esta manera, la idea es generar una evaluación de impacto controlada entre ambos experimentos. Los detalles del método de evaluación pueden ser encontrados en la sección 3.4.2.

- **Estrategias de Castigo**: La presencia de estrategias de castigo entre agentes en contextos de aprendizaje reforzado se revela como un aspecto fundamental para evaluar si estos agentes alcanzan un equilibrio indicativo de colusión. Aunque investigaciones previas han centrado su atención principalmente en los resultados emergentes, es crucial reconocer que la mera presencia de precios elevados no constituye una evidencia concluyente de prácticas colusivas. La colusión económica se caracteriza por un esquema complejo de incentivos y penalizaciones diseñado para estabilizar los precios a niveles consistentemente superiores al umbral competitivo. Por tanto, es esencial diferenciar entre colusión genuina y anomalías en la optimización estratégica. Esta distinción es fundamental para identificar comportamientos que podrían ser considerados preocupa-

ciones antimonopolísticas y para lograr una comprensión exhaustiva de estos fenómenos.

En consonancia con lo anterior, y a base de las metodologías propuestas por [2], [26], [3] y [28], el objetivo de este experimento es evaluar si el equilibrio supracompetitivo alcanzado tiene su origen en acciones cooperativas. Para lograr esto ultimo, se sugiere una evaluación de la existencia de estrategias de castigo mediante la inducción de desviaciones en un agente y el análisis subsiguiente de la respuesta de su contraparte. En particular, se plantea la imposición de una condición donde uno de los agentes adopte el precio correspondiente al equilibrio de Nash en $\bar{t}$, expresado mediante la siguiente ecuación:

$$p_{0,\bar{t}} = p_{N,\bar{t}} \tag{3.21}$$

De esta forma, se evalúa como caso favorable aquellos casos donde el agente responde al desvío con una reducción en su política de precios en los 5 *timestep* siguientes al desvío, y un caso negativo aquellos casos donde el agente no realiza ningún ajuste en sus acciones en la misma ventana de tiempo. Finalmente, considerando que lo deseable es evaluar la existencia de estrategias de castigo una vez los agentes hayan aprendido la estrategia más óptima, se propone forzar el desvío de los agentes en $\bar{t} = 350000$.

- **Sensibilidad de hiperparámetros**: Para realizar un análisis detallado de la sensibilidad de los resultados obtenidos en el experimento base, se procede a la modificación gradual de los siguientes hiperparámetros:
  - Cantidad de períodos pasados k
  - Número de agentes N
  - Probabilidad de shock inflacionario ρ
  - Tasa de aprendizaje lr
  - Factor de descuento γ

El propósito es obtener una comprensión más profunda de la influencia de cada hiperparámetro en la competitividad del mercado. Este análisis se efectúa siguiendo el principio de *ceteris paribus*, es decir, se altera solo el hiperparámetro de interés manteniendo el resto constante, conforme a las condiciones del experimento base. Concretamente, se propone evaluar tres valores posibles para cada hiperparámetro, los cuales son especificados en la siguiente tabla:

Tabla 3.2: Valores de hiperparámetros a experimentar.

Hiperparámetro	**Base**	**Valor 1**	**Valor 2**
Cantidad de períodos pasados (k)	1	10	25
Número de Agentes (N)	2	5	10
Probabilidad de shock inflacionario (ρ)	0.001	0.002	0.003
Tasa de aprendizaje (lr)	0.1	0.2	0.3
Factor de descuento (γ)	0.95	0.8	0.7

- **Agente altruista**: Finalmente, con el fin de explorar posibles vías para mitigar los efectos nocivos de la competencia algorítmica sobre la competitividad de los mercados y en función de lo estipulado en la sección 3.2, se diseña un experimento en el que el agente 0 fije un precio equivalente al equilibrio de Nash, es decir:

$$p_{0,t} = p_{N,t} \tag{3.22}$$

De esta manera, el objetivo es evaluar las rentabilidades obtenidas por los agentes al incorporar este tipo de agente al mercado. Cabe destacar que, aunque este experimento guarda una notable similitud con el experimento de Estrategias de Castigo, se diferencian en la cantidad de *timesteps* durante los cuales se fuerza al Agente 0 a utilizar el precio del equilibrio de Nash.

3.4.2. Evaluación Experimental

- **Evaluación de competitividad**: Para evaluar el nivel de competitividad en el mercado, se adopta la metodología propuesta por [2], en la cual se define Δ_t como un indicador de rentabilidad relativa a los equilibrios de Nash y Monopólico. Esto permite una comparación más coherente de las métricas a través de diferentes periodos temporales:

$$\Delta_t = \frac{\bar{R}_t - R_t^N}{R_t^M - R_t^N} \tag{3.23}$$

Donde $\bar{R}_t$ simboliza la rentabilidad promedio obtenida por los agentes en el instante t, y (R^N, R^M) denota las rentabilidades correspondientes a los equilibrios de Nash y Monopólico, respectivamente.

Considerando que las acciones están ancladas a los costos (ver ecuación 3.10) y los costos pueden no crecer al mismo ritmo que los equilibrios de Nash y Monopolio, las mediciones de Δ_t poseen un **sesgo** inherente a la diferencia en los crecimiento de la inflación y el equilibrios de Nash y Monopolio. Para solucionar este problema, se define ∇_t como:

$$\nabla_t = \frac{\bar{R}_t - R_t^{N_F}}{R_t^{M_F} - R_t^{N_F}} \tag{3.24}$$

donde $R_t^{N_F}$ y $R_t^{M_F}$ representan las rentabilidades de Nash y Monopolio a las que se les impone un crecimiento equivalente a la inflación. De esta manera, ∇_t logra capturar la existencia de estrategias cooperativas controlando por la diferencia en el crecimiento entre las series. Además, con las ecuaciones 3.23 y 3.24 es posible demostrar la siguiente relación:

$$\Delta_t = (\nabla_t - IE) \cdot \beta_t \tag{3.25}$$

donde IE representa la pérdida en la competitividad atribuida a las diferencias de crecimiento entre los costos y los equilibrios equilibrios económicos, y β_t captura la razón de cambio entre Δ_t y ∇_t. Los detalles matemáticos de esta expresión pueden ser encontrados en el Anexo B.1.

Bajo este esquema, valores próximos a 0 de (Δ_t, ∇_t) indican niveles de rentabilidad cercanos a los de un mercado competitivo, en tanto que un valor cercano a 1 señala la existencia de un equilibrio no competitivo y potencialmente dañino para los consumidores. Por otro lado, un Δ_t negativo sugiere que los agentes están adoptando estrategias subóptimas, lo cual implica la posibilidad de alcanzar mayores recompensas ajustándose al precio de equilibrio de Nash, sin comprometer sus niveles de rentabilidad. Finalmente y a modo de lograr una evaluación limpia del proceso de exploración, se implementa la evaluación considerando sólo los últimos 50.000 *timesteps* de cada experimento.

- **Robustez Estadística**: Con tal de conseguir robustez estadística en las conclusiones de los experimentos antes señalados, se busca repetir 50 veces cada uno de estos experimentos. De esta manera, cada experimento aislado se diferencia de otro a partir de:

  - La serie inflacionaria de donde extraer π_t
  - Los pesos de inicialización de las redes neuronales
  - La estrategia escogida por *epsilon-greedy* durante la etapa de exploración

  Por otro lado, y con el objetivo de estimar con mayor exactitud el impacto de las diferentes variaciones señaladas en los experimentos, se busca utilizar un test T de muestras emparejadas. Este test es particularmente útil cuando se comparan dos muestras que están relacionadas o emparejadas de alguna manera, lo que resulta especialmente útil para la evaluación de la sensibilidad de hiperparámetros. La fórmula para el test T de muestras emparejadas es la siguiente:

  $$t = \frac{\overline{d}}{s_d/\sqrt{n}}$$

  donde $\overline{d}$ es la media de las diferencias entre las pares de muestras, s_d es la desviación estándar de estas diferencias y n es el número de pares.

  Además, se utiliza *Cohen's d* [34] como estadístico descriptivo para medir el tamaño del efecto, el cual se calcula como la diferencia entre las medias de dos grupos dividida por la desviación estándar de los datos. La fórmula para el *Cohen's d* es:

  $$d = \frac{\overline{X}_1 - \overline{X}_2}{s_p}$$

  donde $\overline{X}_1$ y $\overline{X}_2$ son las medias de las dos muestras emparejadas y s_p es la desviación estándar agrupada de las muestras. De esta manera y siguiendo la metodología propuesta por los autores del método, se busca cuantificar el tamaño del efecto catalogando el valor de *Cohen's d* por medio de la Tabla 3.3.

- **Separación de Ambientes Train-Test**: Finalmente, con el objetivo de complementar cada experimento, se aborda una crítica clave en el campo de *Algorithmic Collusion*: la limitación de evaluar algoritmos únicamente en el entorno en el que fueron entrenados, lo cual dificulta la generalización de los resultados a situaciones reales de competencia. Para superar esta limitación, se adopta una metodología avanzada inspirada en los

trabajos de [27], que diferencia claramente entre los entornos de entrenamiento y prueba.

El diseño experimental se divide en dos fases: primero, se entrena a dos agentes de manera independiente en entornos separados, según las pautas de un experimento base. Luego, en la fase de prueba, estos agentes compiten en un nuevo entorno con una semilla de inicialización diferente, congelando sus pesos antes de esta etapa para mantener la integridad de sus estrategias. Esta metodología no solo evalúa la efectividad de los algoritmos en un entorno controlado, sino que también ofrece indicios sobre su aplicabilidad en escenarios competitivos reales, brindando una perspectiva más amplia sobre su capacidad de funcionar en diferentes contextos.

Tabla 3.3: Intervalos de d y su correspondiente tamaño del efecto.

d	**Tamaño del Efecto**
$d < 0.01$	Despreciable
$0.01 \leq d < 0.20$	Muy pequeño
$0.20 \leq d < 0.50$	Pequeño
$0.50 \leq d < 0.80$	Medio
$0.80 \leq d < 1.20$	Grande
$1.20 \leq d < 2.00$	Muy grande
$d \geq 2.00$	Enorme

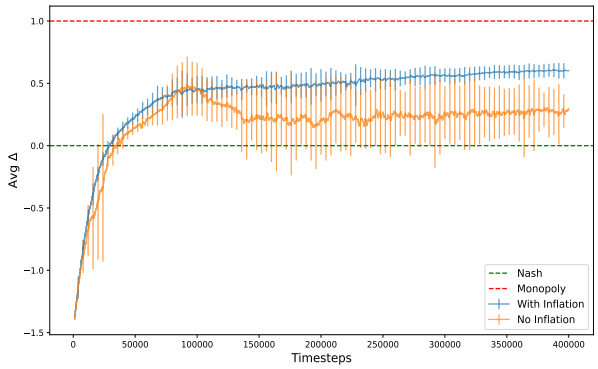
Capítulo 4

Resultados y Análisis

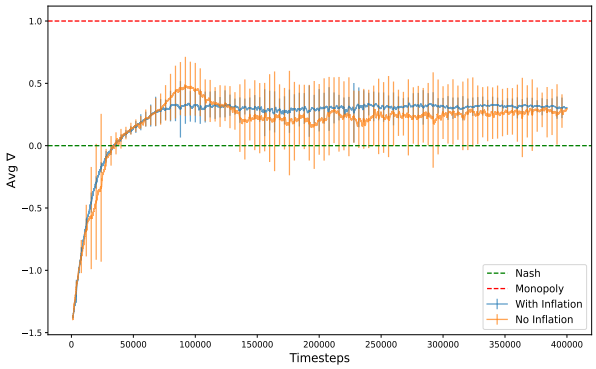
En esta sección, se exponen los resultados obtenidos tras la implementación de los experimentos descritos en la Sección 3.4. La presentación de estos resultados se aborda en función de Δ y ∇, es decir, del nivel de rentabilidades y cooperación que se obtiene a partir de la interacción entre los agentes. Además, se complementa el análisis examinando el margen sobre costos η (ver Ecuación 3.12) escogido por los agentes a lo largo de los experimentos. Para facilitar una visualización más clara de las series temporales, se grafica la media móvil de las series con una ventana temporal de 1000 *timesteps*. Finalmente, junto a cada experimento se adjunta una tabla con estadísticos calculados sobre los últimos 50000 *timesteps*. En concreto, se presenta el promedio μ y desviación σ de Δ y ∇, junto al estadístico d resultante de implementar el Test-T de muestras emparejadas y usando la evaluación *Cohen's d*.

4.1. Experimento base

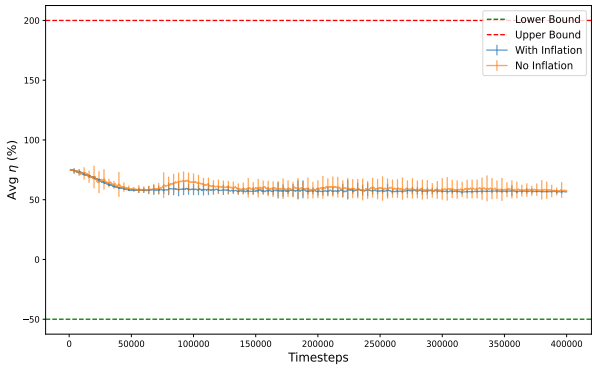
El principal objetivo de este experimento consiste en evaluar la persistencia de los fenómenos identificados en la literatura sobre *Algorithmic Collusion*, particularmente bajo la influencia de la variabilidad en los costos de producción de los agentes. Para tal fin, se establecen dos escenarios distintos: uno sin inflación, donde la probabilidad de ocurrencia de shocks inflacionarios ρ es igual a cero, y otro donde se introducen shocks inflacionarios en los precios, con un ρ establecido en 0.001. Los resultados obtenidos en este experimento se ilustran en la Figura 4.1 y en la Tabla 4.1:



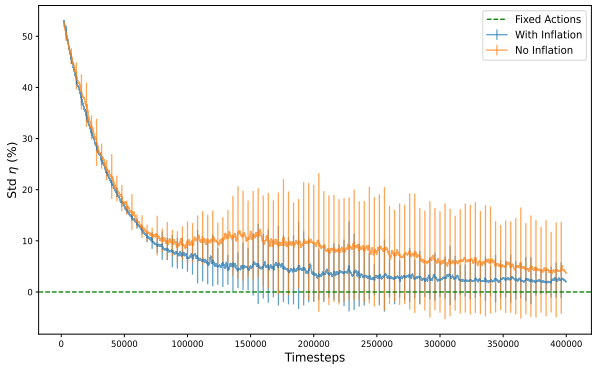
(a) Media móvil de Δ.



(b) Media móvil de ∇.



(c) Media móvil de acciones.



(d) Desviación Estándar Móvil de acciones.

Figura 4.1: Comparativa escenario con y sin inflación.

En función de los resultados obtenidos, se observa como los agentes logran rentabilidades superiores a las del equilibrio de Nash de manera consistente, siendo esto válido para los escenarios con y sin inflación. En particular, se observa que el promedio de Δ llega a 0.5976 para el caso con inflación y 0.2746 para el caso sin inflación, posicionando el tamaño del efecto como "Enorme". Las conclusiones son similares al analizar ∇, donde se observa que el escenario con inflación asciende a un promedio de 0.3156, mientras que el escenario sin inflación llega a 0.2746. De esta manera, el tamaño del efecto de la inflación sobre ∇ es "Pequeño". Este último resultado es interesante, pues arroja indicios de que cambios en la variabilidad de los costos podría afectar de manera positiva a las rentabilidades obtenidas por los agentes. Por otro lado, se observa como el margen cobrado por los agentes (acciones) se mantiene equivalente para ambos escenarios. Si se analizan estos resultados de manera conjunta, se observa que si bien el escenario con inflación genera grandes granancias en rentabilidades, el grueso de estas rentabilidades no son generadas a partir de un configuración de estrategias diferentes, sino que se deben a que los precios están anclados a los costos.

Tabla 4.1: Resultados Experimento Base en configuración de entrenamiento.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
Sin inflación	50	0.2746	0.1664	2.5042	0.2746	0.1664	0.3272
Con inflación	50	0.5976	0.0520	-	0.3156	0.0548	-

En el entorno de prueba, se observa un escenario similar (ver Tabla 4.2): El promedio de

Δ supera el equilibrio de Nash en los casos con y sin inflación. Este resultado es interesante, pues sugiere que los agentes son capaces de alcanzar escenarios anticompetitivos incluso fuera de la muestra de entrenamiento. Por otro lado, se observa que ocurre un efecto similar al entorno de entrenamiento: el nivel de Δ promedio del escenario con inflación es muy superior a aquel que no considera inflación. En específico, el escenario sin inflación obtiene un promedio de Δ igual a 0.2920, mientras que el escenario con inflación obtiene un Δ igual a 0.3678, clasificando de esta manera como "Medio" al tamaño de impacto de considerar inflación. Nuevamente, esto se puede explicar en que las acciones de los agentes están ancladas a los costos. Del mismo modo, la diferencia en ∇ se reduce bastante, catalogando el tamaño del efecto como "Pequeño".

Tabla 4.2: Resultados Experimento Base en configuración de prueba.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
Sin inflación	50	0.2920	0.1405	0.6679	0.2920	0.1405	0.3047
Con inflación	50	0.3678	0.0671	-	0.3258	0.0639	-

Este aumento en el valor de ∇, aunque es apenas apreciable, es de especial relevancia, ya que sugiere que un incremento en los costos de producción podría reducir la competitividad en los mercados. Tal como lo plantea [35], una explicación plausible es que la inflación crea un ambiente de incertidumbre, desincentivando la competencia agresiva en precios y fomentando, en cambio, un consenso tácito hacia niveles de precios más elevados. Otra explicación posible para explicar este fenómeno radica en que el aumento en los costos de producción podría actuar como un mecanismo de coordinación implícita [36]. Bajo esta perspectiva, los agentes, enfrentando desafíos similares, podrían tender a adoptar estrategias de precios más alineadas. Estos resultados no solo validan las hipótesis iniciales del estudio, sino que también abren caminos para futuras investigaciones sobre el impacto de las variaciones en los costos de producción en la estructura competitiva de los mercados.

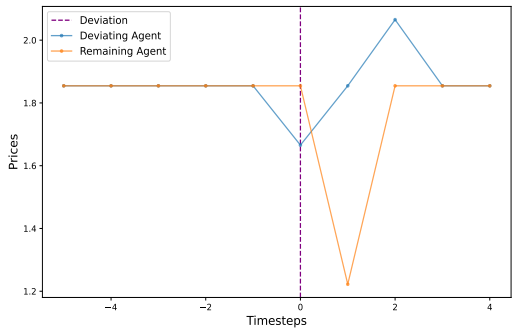
4.2. Estrategia de Castigo

De acuerdo con varios autores en la literatura sobre *Algorithmic Collusion*, la presencia de estrategias de castigo entre agentes es un indicativo de la existencia de estrategias cooperativas. Para probar la presencia de tales estrategias, se entrenan los agentes siguiendo una configuración base y, posteriormente, se fuerza a uno de los agentes a ajustar su precio al equilibrio de Nash en la última fracción del experimento, conforme a la ecuación 3.21. De esta forma, el objetivo es analizar la respuesta del agente restante mediante el monitoreo de su política de precios, prestando especial atención a los casos en que dicho agente responde al desvío con una reducción en sus precios. La Tabla 4.3 presenta los resultados de este experimento.

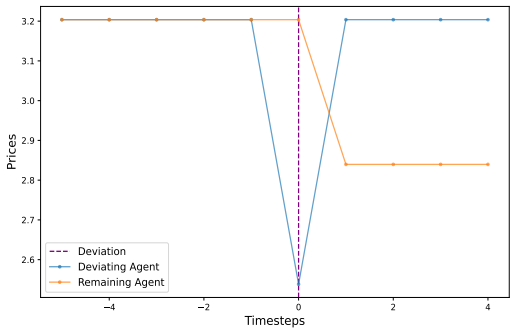
Tabla 4.3: Resultados de validación de estrategias de castigo.

N	**Casos con castigo**	**Casos sin castigo**
50	2	48

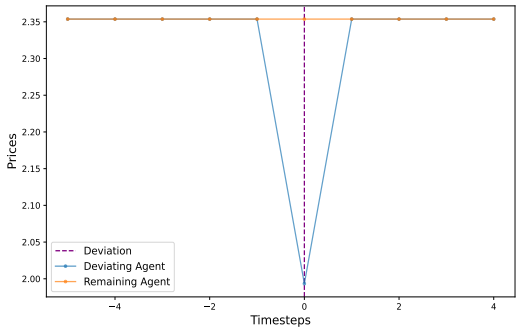
Los resultados indican que, de los 50 experimentos ejecutados, sólo en 2 de ellos se encuentra un cambio en la política de precios como posible castigo, mientras que para el resto de experimentos el agente es incapaz de ejecutar una respuesta acorde. La Figura 4.2 ilustra algunos ejemplos de este experimento.



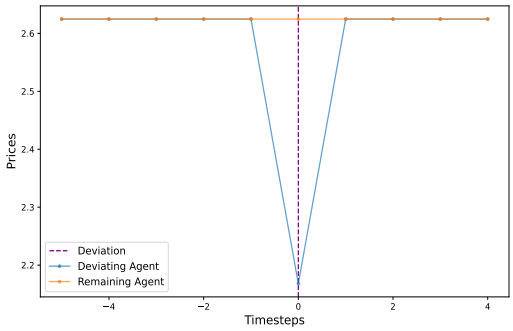
(a) Caso favorable: Experimento 43.



(b) Caso favorable: Experimento 45.



(c) Caso desfavorable: Experimento 17.



(d) Caso desfavorable: Experimento 25.

Figura 4.2: Resultados evaluación estrategias de castigo.

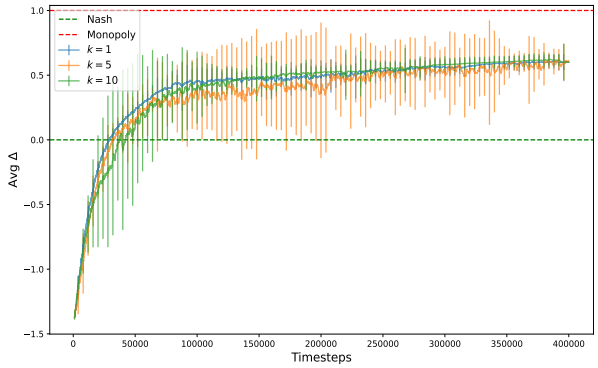
Este patrón en los resultados es notable, ya que sugiere que las rentabilidades monopólicas observadas pueden no derivarse de una estrategia cooperativa entre los agentes, sino mas bien de que los agentes fueron incapaces de generar una política óptima para resolver el Proceso de Decisión de Markov, fijando precios por encima del equilibrio de Nash y dando espacio a la confusión con comportamientos colusivos. Este resultado hace sentido con los estudios de [27] y [28], donde se propone que los resultados supra competitivos de los estudios de *Algorithmic Collusion* no obedecen a estrategias de colusión. Considerando la reducida tasa de éxito en donde los agentes logran adaptarse (tan solo 4 % del total de experimentos), se concluye que no se cuenta con la evidencia necesaria para demostrar la existencia de estrategias de castigo entre los agentes.

4.3. Sensibilidad a hiperparámetros

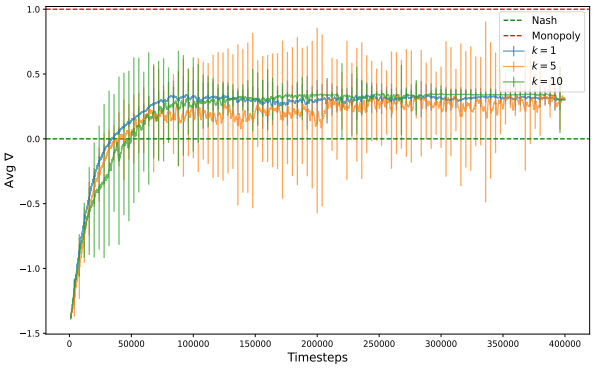
Acorde a lo especificado en la sección 3.4, esta sección presenta los resultados de variar gradualmente cada hiperparámetro de interés en base a lo expresado en la tabla 3.2. A continuación, se exponen los resultados obtenidos de estos experimentos detallados.

4.3.1. Cantidad de períodos pasados k

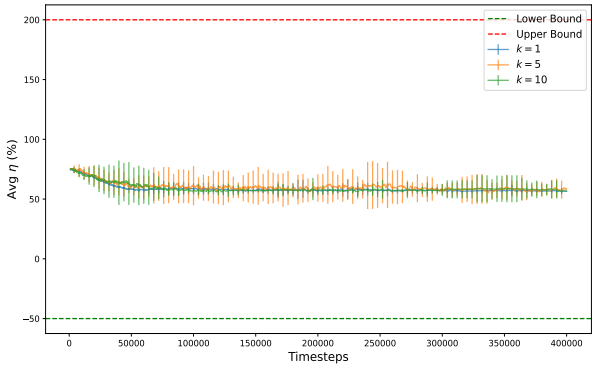
En la presente sección, se aborda un análisis exhaustivo para evaluar la sensibilidad de los resultados previamente obtenidos ante variaciones en el número de períodos pasados disponibles (k) para la observación de los agentes económicos. Con el objetivo de discernir el impacto de diferentes volúmenes de información histórica en la toma de decisiones de los agentes, se lleva a cabo una serie de experimentaciones utilizando tres valores distintos para k: 1, 10 y 25. La Figura 4.3 y tabla 4.4 ilustran los resultados de estos experimentos:



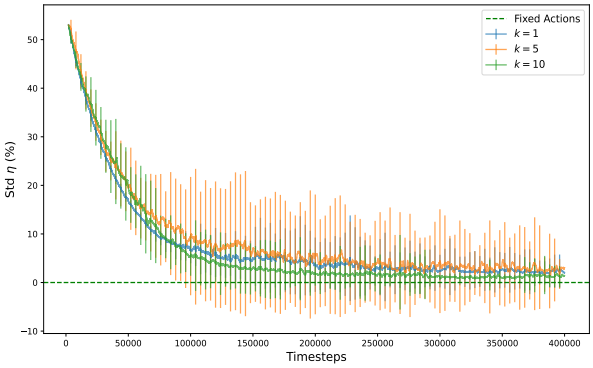
(a) Media móvil de Δ.



(b) Media móvil de ∇.



(c) Media móvil de acciones.



(d) Desviación Estándar Móvil de acciones.

Figura 4.3: Resultados de sensibilidad - Cantidad de períodos pasados k.

El análisis de los datos recabados revela resultados significativos en términos de rentabilidad y estabilidad del equilibrio en el mercado. Se constata que para todas las combinaciones del parámetro k, que representa el número de períodos pasados observados, se alcanzan rentabilidades superiores al *Equilibrio de Nash*. En particular, se observa que cuando $k = 10$ el promedio de Δ cae a 0.5862 (efecto "Muy pequeño"), mientras que si $k = 25$ sube a 0.6106 (efecto "Medio"). Este resultado es contraintuitivo, pues indica que un aumento en k no siempre termina en un aumento en las rentabilidades de los agentes. En relación a ∇, cuando $k = 10$ el promedio cae a 0.2964 (efecto "Muy pequeño"), mientras que si $k = 25$ sube a 0.3370 (efecto "Medio").

Una conclusión similar se puede extraer a partir de los resultados en en el entorno de prueba (ver Tabla 4.5). Si bien se aprecian niveles de rentabilidad superiores al equilibrio de Nash, se obtienen conclusiones contradictorias con respecto a los resultados obtenidos en el entorno de entrenamiento en relación al número de períodos k. En específico, se observa que

Tabla 4.4: Resultados del experimento para diferentes configuraciones de k en configuración de entrenamiento.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$k = 1$ (Base)	50	0.5976	0.0520	-	0.3156	0.0548	-
$k = 10$	50	0.5862	0.1164	0.1492	0.2964	0.1871	0.1467
$k = 25$	50	0.6106	0.0453	0.5323	0.3370	0.0170	0.5335

cuando $k = 10$, los niveles promedio de Δ caen a 0.1320 (efecto "Pequeño"), mientras que si $k = 25$ este cae a 0.3550 (efecto "Pequeño").

Tabla 4.5: Resultados del experimento para diferentes configuraciones de k en entorno de prueba.

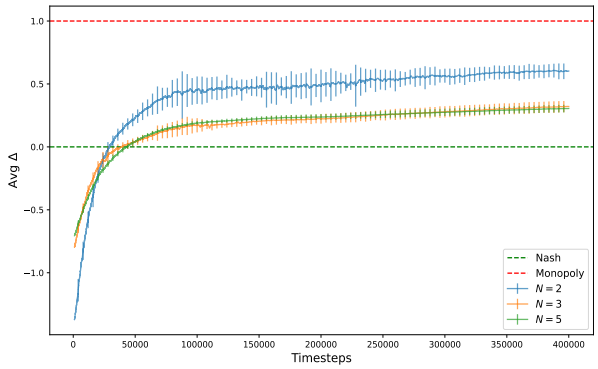
Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$k = 1$ (Base)	50	0.3678	0.0671	-	0.3258	0.0639	-
$k = 10$	50	0.1320	1.0029	0.3326	0.0818	1.0306	0.3330
$k = 25$	50	0.3550	0.0919	0.1574	0.3118	0.1012	0.1609

En retrospectiva de los resultados obtenidos a lo largo de esta sección, se concluye que se necesita de más evidencia para poder realizar aseveraciones con respecto al impacto de este parámetro sobre los niveles de competitividad en el mercado. El hecho de que el sentido del efecto sobre Δ varíe a mayores niveles de k va en disonancia con los hallazgos reportado por Calvano et al. [2], estando en contra con la lógica económica que sugiere que proporcionar una mayor cantidad de información histórica a los agentes de mercado facilita la coordinación entre ellos. Una posible explicación a estos resultados puede residir en la falta de estacionariedad en el ambiente, causada principalmente por su propiedad multiagente y los shocks inflacionarios. De esta forma, el no cumplimiento de esta propiedad podría impactar de forma negativa en la convergencia de los agentes, generando comportamiento sub-óptimos con un mayor número de períodos k.

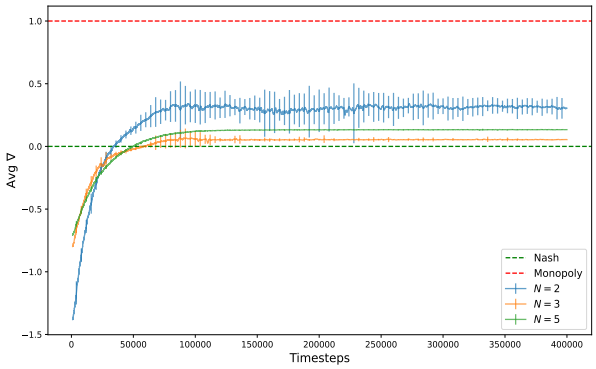
4.3.2. Número de agentes N

En esta sección de la tesis, se realiza un análisis detallado para determinar cómo la variabilidad en el número de agentes económicos involucrados en el experimento afecta a los resultados previamente obtenidos. Este análisis se enfoca en el parámetro N, que representa el número de agentes activos en el experimento. Para comprender el impacto de un incremento en la cantidad de agentes, se han diseñado y ejecutado una serie de experimentos utilizando tres configuraciones distintas para N: 2, 5 y 10.

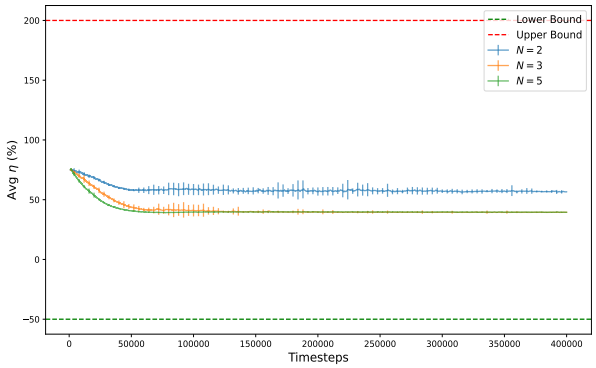
La Figura 4.4 y Tabla 4.6 presentan los detalles de este experimento en el entorno de entrenamiento.



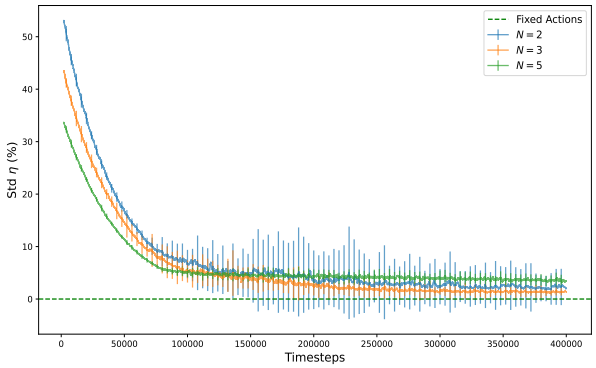
(a) Media móvil de Δ.



(b) Media móvil de ∇.



(c) Media móvil de acciones.



(d) Desviación Estándar Móvil de acciones.

Figura 4.4: Resultados de sensibilidad - Número de agentes N.

Al examinar los resultados obtenidos, se observa como el número de agentes logra impactar de manera negativa a las rentabilidades de manera efectiva. Específicamente, cuando $N = 3$ se observa que el nivel promedio de Δ cae a 0.3132 (efecto "Enorme"), mientras que si $N = 5$ este nivel cae a 0.2986 (efecto "Enorme"). De manera similar, al analizar los niveles de ∇ se obtienen conclusiones similares, obteniendo un promedio de ∇ igual a 0.0506 cuando $N = 3$ y 0.1300 cuando $N = 5$ (tamaños de efecto "Enorme" en ambos casos). Este último hallazgo es interesante, pues implica que un mayor número de agentes en el experimento afecta de forma negativa a las rentabilidades de los agentes, llegando incluso a equilibrios marginalmente diferentes al equilibrio de Nash en el caso de ∇. Es interesante notar también la caída en la varianza de ∇ con un mayor número de agentes, lo que se puede explicar en la convergencia hacia el equilibrio de Nash. Por otro lado, si bien los niveles de Δ caen fruto de un mayor número de agentes, la caída es amortiguada por el anclamiento del precio a las costos c_t.

Tabla 4.6: Resultados del experimento para diferentes configuraciones de N en configuración de entrenamiento.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$N = 2$ (Base)	50	0.5976	0.0520	-	0.3156	0.0548	-
$N = 3$	50	0.3132	0.0440	11.5487	0.0506	0.0042	6.8551
$N = 5$	50	0.2986	0.0301	11.8446	0.1300	0.0000	4.7916

En los resultados en el entorno de prueba (ver Tabla 4.7), se obtienen resultados similares

en todas sus configuraciones. En particular, se observan caídas en los niveles promedio de Δ tanto para $N = 3$ y $N = 5$, siendo catalogadas como efecto "Enorme" en ambos casos. Un caso parecido se observa en función de ∇, encontrando caídas catalogadas como efecto "Enorme" en ambos escenarios.

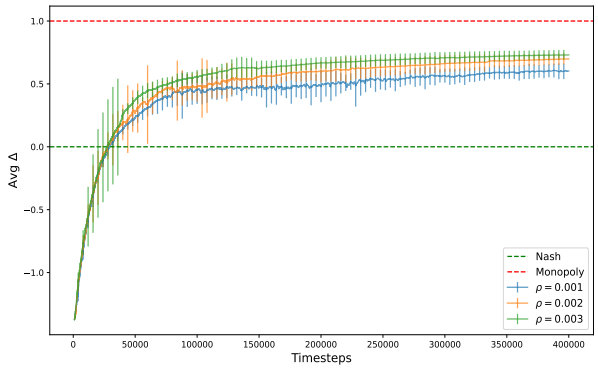
Tabla 4.7: Resultados del experimento para diferentes configuraciones de N en configuración de prueba.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$N = 2$ (Base)	50	0.3678	0.0671	-	0.3258	0.0639	-
$N = 3$	50	0.0900	0.0200	6.1856	0.0520	0.0127	5.9809
$N = 5$	50	0.1594	0.0284	3.9890	0.1422	0.0303	3.5270

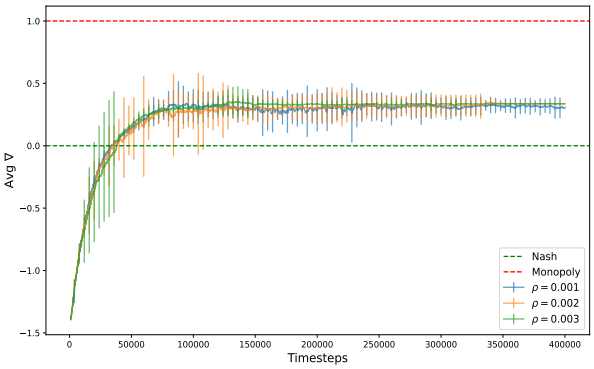
En consideración de los resultados obtenidos, se constata que una mayor cantidad de agentes conduce a una reducción en la dispersión de las rentabilidades obtenidas. Esta observación es congruente con las expectativas teóricas y se alinea con principios económicos fundamentales. Según [37], un incremento en el número de agentes en el mercado conlleva a una intensificación de la competitividad, lo que suele acercar los resultados al equilibrio de Nash, como planteado en la teoría de juegos por [38]. Esta mayor competitividad, como consecuencia, promueve una mayor estabilidad en la política de precios adoptada por los agentes, un fenómeno respaldado por los estudios de [39]. Estas teorías proporcionan un marco robusto para comprender las dinámicas observadas en el experimento. Es importante señalar que ambos conjuntos de resultados, tanto para los escenarios con menor como con mayor número de agentes, presentan diferencias significativas en comparación con el valor base, alcanzando un nivel de confianza del 99 %.

4.3.3. Probabilidad de shock inflacionario ρ

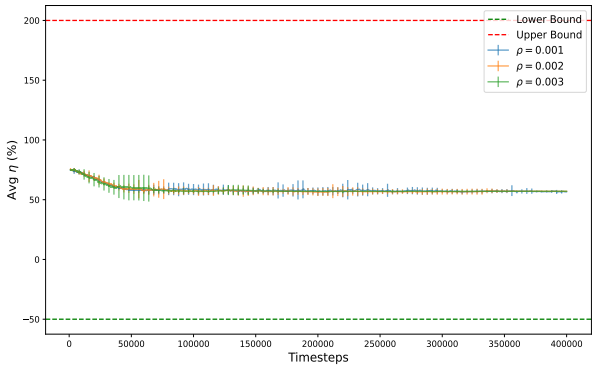
En esta sección del artículo, se profundiza en los experimentos con el fin de cuantificar la sensibilidad de la competitividad del mercado en estudio ante variaciones en la probabilidad de shock inflacionario, denotada por ρ. Se experimentan con tres valores posibles de ρ: 0.001, 0.002 y 0.003. Tal como se discute en la sección 3.4, se controla para que las serie inflacionaria sea la misma para los tres escenarios..La Figura 4.5 y la Tabla 4.8 presentan los resultados de haber implementado este ejercicio.



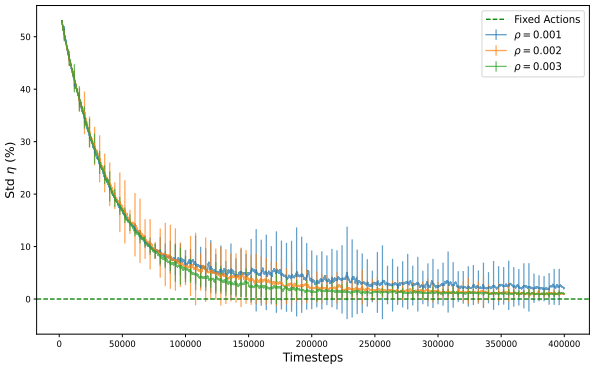
(a) Media móvil de Δ.



(b) Media móvil de ∇.



(c) Media móvil de acciones.



(d) Desviación Estándar Móvil de acciones.

Figura 4.5: Resultados de sensibilidad - Prob. de shock inflacionario ρ.

En función de los resultados y al igual que en los otros experimentos, se observa como los agentes alcanzan rentabilidades sobre las competitivas en los 3 casos. Además, se observa como una mayor probabilidad en el shock inflacionario impacta de manera positiva en el promedio de Δ, catalogando el impacto como "Enorme" para ambos casos. Este resultado es interesante, ya que muestra evidencia que podría apoyar que un mayor nivel de inflación podría terminar en menores niveles de competitividad, apoyando lo encontrado en [40] y [37]. Analizando los resultados encontrados al usar ∇, se observa como el impacto en la probabilidad de inflación es apenas apreciable sobre las acciones de los agentes, obteniendo una categoría de impacto "Medio" en ambos casos. De esta forma, se repiten las mismas conclusiones que en los otros experimentos: Si bien la probabilidad de shock inflacionario parece tener un impacto enorme sobre las rentabilidades de los agentes, el grueso de estas rentabilidades son explicadas en que los precios están anclados en los costos y no a cambios en la estrategia de los agentes.

Tabla 4.8: Resultados del experimento para diferentes configuraciones de ρ en configuración de entrenamiento.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$\rho = 0.001$ (Base)	50	0.5976	0.0520	-	0.3156	0.0548	-
$\rho = 0.002$	50	0.6912	0.0471	3.7347	0.3386	0.0040	0.5859
$\rho = 0.003$	50	0.7258	0.0407	5.2011	0.3388	0.0059	0.5967

Un escenario similar se aprecia al analizar los resultados en el entorno de prueba. Si bien se aprecian rentabilidades sobre las competitivas para todos los casos, el impacto de la probabilidad de shock inflacionario es limitado. En particular, cuando $\rho = 0.002$, se obtiene un impacto catalogado sobre Δ como "Pequeño", mientras que si $\rho = 0.003$ el impacto es "Mediano". En cuanto a ∇, se obtienen impactos catalogados como "muy pequeño" para ambos casos. De esta manera, se concluye que, bajo la formulación experimental propuesta, escenarios de prueba con un mayor nivel de inflación no tienen impacto en los niveles de competitividad del mercado.

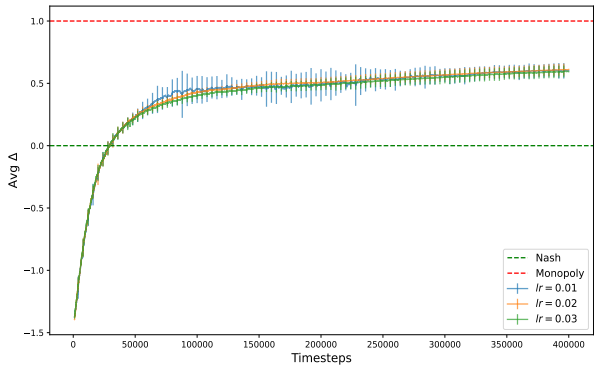
Tabla 4.9: Resultados del experimento para diferentes configuraciones de ρ en configuración de prueba.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$\rho = 0.001$ (Base)	50	0.3678	0.0671	-	0.3258	0.0639	-
$\rho = 0.002$	50	0.3944	0.0799	0.3753	0.3222	0.0880	0.0458
$\rho = 0.003$	50	0.4174	0.0801	0.7154	0.3228	0.0851	0.0390

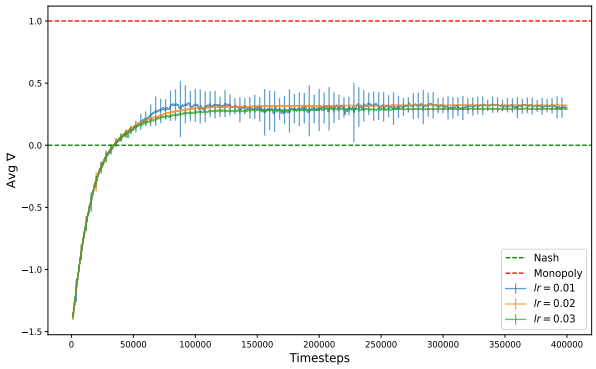
4.3.4. Tasa de aprendizaje lr

En esta sección del artículo, se profundiza en los experimentos con el fin de cuantificar la sensibilidad de la competitividad del mercado en estudio ante variaciones en la tasa de aprendizaje de los agentes, denotada por lr. Se experimenta con tres valores posibles de lr: 0.01, 0.02 y 0.03.

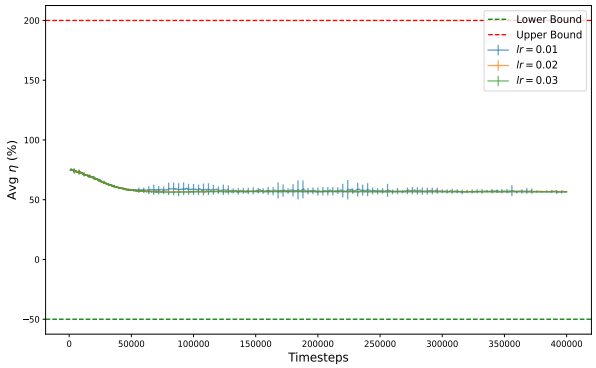
La Figura 4.6 ofrece una representación visual de los resultados obtenidos en estos experimentos. Complementando la información gráfica, las Tablas 4.10 y 4.11 proporcionan un resumen consolidado de los resultados de este experimento en las configuraciones de entrenamiento y prueba, respectivamente.



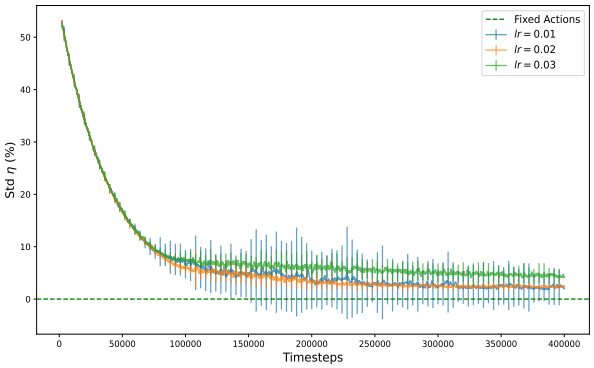
(a) Media móvil de Δ.



(b) Media móvil de ∇.



(c) Media móvil de acciones.



(d) Desviación Estándar Móvil de acciones.

Figura 4.6: Resultados de sensibilidad - Tasa de aprendizaje lr.

Los resultados obtenidos indican que la tasa de aprendizaje (lr) ejerce un impacto mixto sobre las métricas evaluadas. Específicamente, se observa que con $lr = 0.02$, el valor de Δ crece a 0.6022 (tamaño de efecto "Muy pequeño"). Por otro lado, cuando $lr = 0.03$, Δ disminuye cae a 0.5868 (una reducción de 0.01 respecto a la configuración base y teniendo un impacto "Pequeño"). De manera similar, los valores de ∇ suben cuando $lr = 0.02$, pero caen cuando $lr = 0.03$ (con tamaños de efecto "Medio" y "Pequeño", respectivamente). Estos resultados sugieren que un incremento en la tasa de aprendizaje puede conducir a una mayor variabilidad en las acciones de los agentes, resultando en una disminución de la estabilidad de los resultados.

Tabla 4.10: Resultados del experimento para diferentes configuraciones de lr en configuración de entrenamiento.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$lr = 0.01$ (Base)	50	0.5976	0.0520	-	0.3156	0.0548	-
$lr = 0.02$	50	0.6022	0.0464	0.1845	0.3202	0.0014	0.5969
$lr = 0.03$	50	0.5868	0.0479	0.4463	0.2940	0.0063	0.1175

Los resultados en la configuración de prueba ofrecen conclusiones similares (ver Tabla 4.11). Cuando lr aumenta a 0.02, el nivel promedio de Δ sube a 0.3718 (efecto "Muy pequeño"), mientras que si $lr = 0.03$ este cae a 0.3114 (efecto "Pequeño"). Conclusiones parecidas

se obtienen de analizar ∇ para este experimento, obteniendo un tamaño de efecto "Muy pequeño" cuando $lr = 0.02$ y "Pequeño" cuando $lr = 0.03$.

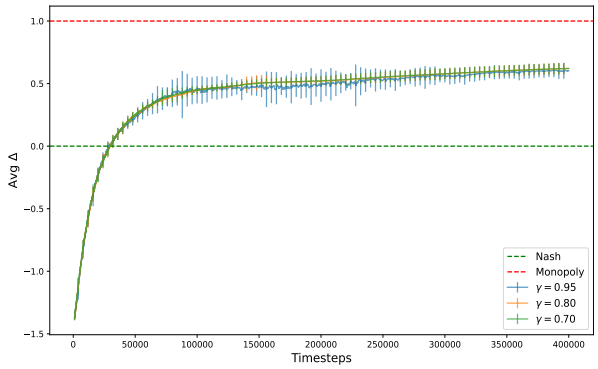
Tabla 4.11: Resultados del experimento para diferentes configuraciones de lr en configuración de prueba.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$lr = 0.01$ (Base)	50	0.3678	0.0671	-	0.3258	0.0639	-
$lr = 0.02$	50	0.3718	0.0867	0.0527	0.3296	0.0928	0.0470
$lr = 0.03$	50	0.3114	0.1729	0.4857	0.2670	0.1712	0.4862

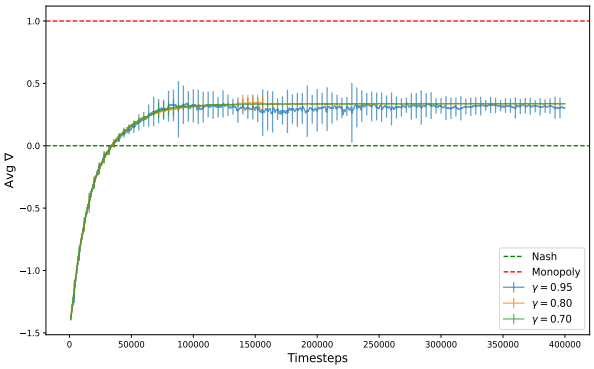
Una explicación plausible para este fenómeno es que los agentes aún no hayan convergido hacia la política óptima para resolver el ambiente, lo que se traduce en una mayor varianza en sus acciones. Esta hipótesis cobra relevancia al considerar que el algoritmo de Deep Q-Networks (DQN) es altamente sensible a los hiperparámetros establecidos, pudiendo variar significativamente los resultados entre diferentes configuraciones [41]. Finalmente, se determina que tanto el incremento como la disminución observados en Δ son estadísticamente significativos y difieren del valor base con un 99 % de confianza. Esta significancia estadística refuerza la importancia de la selección adecuada de la tasa de aprendizaje en la configuración de algoritmos de aprendizaje reforzado.

4.3.5. Factor de descuento γ

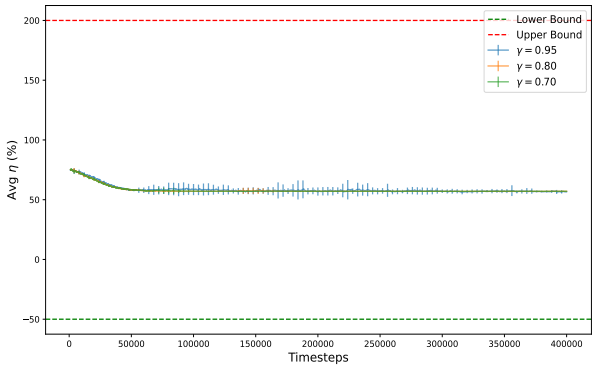
En esta sección del artículo, se profundiza en los experimentos con el fin de cuantificar la sensibilidad de la competitividad del mercado en estudio ante variaciones en la tasa de descuento intertemporal de los agentes, denotada por γ. Se experimenta con tres valores posibles de γ: 0.95, 0.8 y 0.7. La Figura 4.7 y Tabla 4.12 contienen los resultados de haber implementado este experimento en el entorno de entrenamiento.



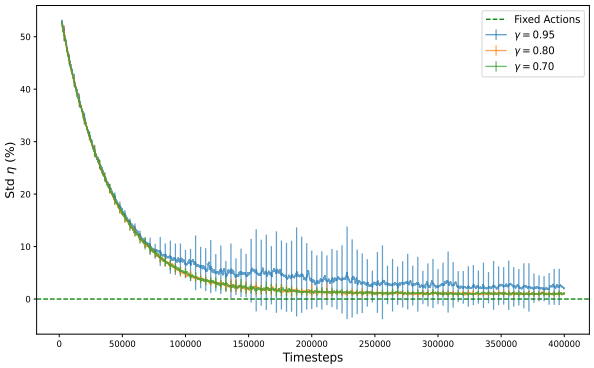
(a) Media móvil de Δ.



(b) Media móvil de ∇.



(c) Media móvil de acciones.



(d) Desviación Estándar Móvil de acciones.

Figura 4.7: Resultados de sensibilidad - Tasa de aprendizaje γ.

Los resultados obtenidos revelan una tendencia consistente hacia rentabilidades que superan los niveles competitivos en todos los experimentos realizados. Esta observación sugiere una correlación positiva entre las rentabilidades monopólicas y la magnitud del factor de descuento intertemporal aplicada. Específicamente, se identifica que con un valor de $\gamma = 0.80$, el parámetro Δ aumenta a 0.6116, mientras que si $\gamma = 0.70$ el valor de Δ se posiciona en 0.6128 (efectos catalogados como de tamaño "Medio" para ambos casos). El relato se repite en torno a ∇, donde una menor tasa de descuento intertemporal impacta de manera positiva en las rentabilidades obtenidas por los agentes, obteniendo impactos catalogados de tamaño "Medio" en ambos casos.

Tabla 4.12: Resultados del experimento para diferentes configuraciones de γ en configuración de entrenamiento.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$\gamma = 0.95$ (Base)	50	0.5976	0.0520	-	0.3156	0.0548	-
$\gamma = 0.8$	50	0.6116	0.0448	0.5805	0.3392	0.0027	0.6119
$\gamma = 0.7$	50	0.6128	0.0458	0.6150	0.3394	0.0023	0.6112

La Tabla 4.13 presenta los resultados de implementar este experimento en el entorno de prueba. Similar lo expuesto anteriormente, se observa que a una menor tasa de descuento intertemporal, los niveles promedio de Δ aumentan aunque en una menor proporción (tamaños de efecto catalogados como "Pequeño" y "Muy pequeño" para $\gamma = 0.8$ y $\gamma = 0.7$, respectiva-

mente). Un caso similar se observa en torno a ∇, donde si bien esta métrica aumenta a una menor tasa de descuento intertemporal, el tamaño del impacto es reducido.

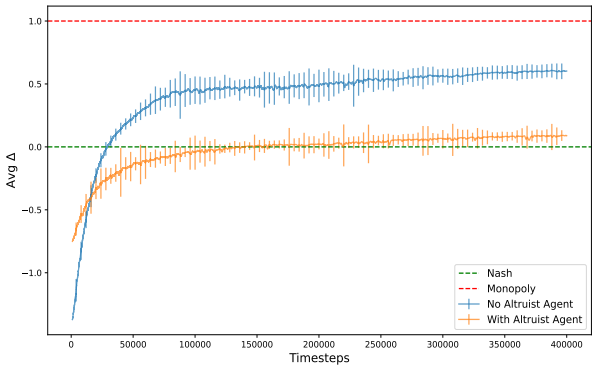
Tabla 4.13: Resultados del experimento para diferentes configuraciones de γ en configuración de prueba.

Configuración	**N**	Δ			∇		
		μ	σ	d	μ	σ	d
$\gamma = 0.95$ (Base)	50	0.3678	0.0671	-	0.3258	0.0639	-
$\gamma = 0.8$	50	0.3814	0.0213	0.3124	0.3396	0.0019	0.3047
$\gamma = 0.7$	50	0.3734	0.0593	0.0933	0.3312	0.0607	0.0851

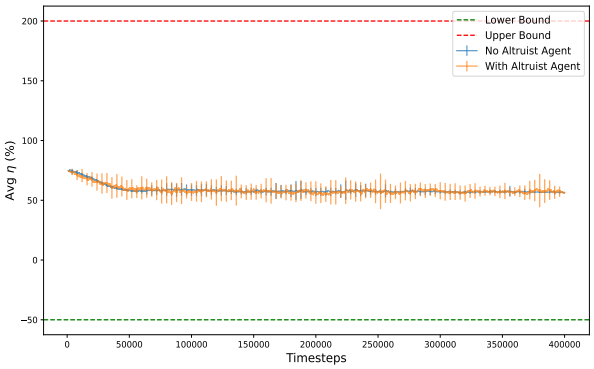
A pesar del impacto reducido encontrado, los resultados presentados van en contra de lo señalado en [2], donde se argumenta que una menor tasa de descuento intertemporal desincentiva estrategias de largo plazo, otorgando mayor importancia a la recompensa por desviarse de estrategias cooperativas teniendo como consecuencia un nivel más cercano al *Equilibrio de Nash*. Al igual que en los experimentos anteriores, una posible explicación a este suceso es la no estacionariedad del ambiente, lo que puede estar generando políticas de acciones en los agentes que disten de ser las óptimas.

4.4. Agente altruista

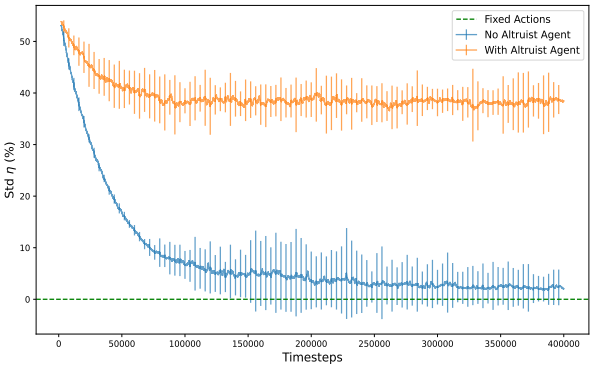
En esta sección del artículo, se profundiza en los experimentos con el fin de cuantificar la sensibilidad de la competitividad del mercado en estudio con la participación de un agente altruista (Agente 0). Considerando que el agente altruista usa el equilibrio de Nash teórico p_N, en esta sección sólo se presentan resultados en torno a Δ. La Figura 4.8 y Tabla 4.14 contienen los detalles de implementar este experimento en el entorno de entrenamiento.



(a) Media móvil de Δ.



(b) Media móvil de acciones.



(c) Desviación estándar móvil de acciones.

Figura 4.8: Resultados de experimento con agente altruista.

Los resultados indican que el agente altruista es efectivo en reducir los niveles de la rentabilidad de las empresas. Específicamente, se destaca una reducción en el nivel promedio de Δ a 0.0834 catalogando el impacto del agente altruista como "Enorme". Esta variación, con una desviación estándar de 0.05, sugiere un equilibrio marginalmente diferente al *Equilibrio de Nash*. Este hallazgo es relevante, ya que demuestra que la inclusión de un agente altruista en el mercado constituye una estrategia efectiva y estadísticamente significativa para disminuir las rentabilidades monopólicas, alterando así la dinámica de mercado y fomentando una competencia más equitativa.

Tabla 4.14: Resultados Experimento Agente Altruista en configuración de entrenamiento.

Configuración	**N**	Δ		
		μ	σ	d
Con agente altruista	50	0.0834	0.0595	11.0864
Sin agente altruista	50	0.5976	0.0520	-

Un escenario similar se observa en el caso de el entorno de prueba, donde el impacto de la presencia del agente altruista afecta negativamente a las rentabilidades obtenidas por los agentes, llegando a un equilibrio marginalmente distinto al *Equilibrio de Nash*. En particular, se observa que la presencia del agente altruista reduce el nivel promedio de Δ a -0.2332, categorizando el tamaño del efecto como "Enorme".

Tabla 4.15: Resultados Experimento Agente Altruista en configuración de prueba.

Configuración	**N**	Δ		
		μ	σ	d
Con agente altruista	50	-0.2332	0.0560	11.0778
Sin agente altruista	50	0.3258	0.0639	-

4.5. Resumen y Discusión de Resultados

El propósito de esta sección es otorgar un resumen y discusión de los resultados encontrados a lo largo de esta investigación. Las Tablas 4.16 y 4.17 el resumen de los resultados encontrados.

El principal propósito de esta investigación es explorar los posibles efectos de incluir variabilidad en los costos sobre los resultados encontrados en la literatura *Algorithmic Collusion*. A través del experimento base, fue posible comprobar como la inclusión (o exclusión) de inflación en la configuración del experimento tiene un impacto significativo, aumentando las rentabilidades obtenidas por los agentes tanto en el ambiente de entrenamiento como en el de prueba. Este resultado es relevante, pues arroja indicios de que la competitividad en los mercados podría verse afectada de forma negativa al participar agentes basados en inteligencia artificial y bajo un escenario fuerte de inflación, factores que han estado cada vez más presentes en el último tiempo. Una segunda mirada en torno a ∇ nos ofrece un mejor entendimiento de este fenómeno: si bien los niveles de competitividad son impactados fuertemente, el grueso de este impacto no se debe a cambio en la estrategias de los agentes, sino mas bien al anclaje de los precios sobre los costos. Si bien este resultado puede parecer tranquilizador, se debe tener especial cuidado con la generación de los precios desde los agentes: basta con que los agentes fijen precios sobre los costos para llegar a los niveles de competitividad obtenidos en el experimento.

En cuanto a la evaluación de estrategias de castigo, los resultados concluyen que no se dispone de evidencia suficiente para demostrar que los agentes ejecutan estrategias de castigo para evitar el desvío del equilibrio. En particular, sólo en 2 de los 50 experimentos ejecutados se observa como los agentes son capaces de responder de manera acorde al desvío del agente. Este resultado va en total disonancia con lo encontrado por [2], donde demuestran que los agentes son capaces de responder a estos desvíos en más del 90 % de las ocasiones. Si bien este resultado podría parecer desalentador, una posible explicación a este fenómeno es la diferencia fundamental en el algoritmo empleado: mientras que este estudio se basa en el algoritmo de aprendizaje reforzado profundo DQN, el estudio de Calvano (y la mayoría de los estudios de *Algorithmic Collusion*) se basan en Q-learning. La diferencia entre ambos algoritmos es sustancial: mientras que Q-learning se basa en una estructura tabular para actualizar el valor de cada par (s, a), DQN usa una red neuronal para estimar el valor de $Q(s, a)$. De esta manera, DQN puede verse como un algoritmo que tiene tanto una mayor capacidad predictiva como una mayor complejidad, lo que puede explicar la diferencia en las estrategias aprendidas por los agentes. De ser válido lo anterior, esto podría indicar que la estrategia de castigo puede no ser universal para todas las configuraciones, limitando en parte el alcance de los resultados expuestos en [2]. Se dispone como trabajo futuro compro-

bar de manera empírica el alcance este tipo de estrategias en función del algoritmo empleado.

Los resultados de implementar el análisis de sensibilidad indican que la configuración es altamente relevante sobre los niveles de competitividad en los mercados. En particular, se observa que la probabilidad de ocurrencia del shock inflacionario ρ tiene un impacto enorme sobre las rentabilidades de los agentes, golpeando fuertemente el nivel de competitividad del mercado. Al igual que en el experimento base, este efecto se puede explicar en función del anclaje de las acciones sobre los costos. Al mirar los resultados de ∇, se observa como el cambio en la estrategia de los agentes es limitado. Este resultado es interesante, pues arroja indicios de que cambios en la configuración tienen un impacto reducido sobre el equilibrio alcanzado entre los agentes sobre el equilibrio base. De esta manera, si se quisiera cambiar los hiperparámetros de los agentes para reducir las rentabilidades de los agentes, estos dificilmente podrían tener un impacto elevado sobre las acciones de los agentes, a menos que se altere de forma estructural las dinámicas del mercado de estudio. Un ejemplo de lo anterior es el número de agentes N del experimento, única variable que fue capaz de generar un impacto significativo sobre las rentabilidades y comportamiento de los agentes para todos los análisis. En específico, se observa que un mayor número de agentes impacta de manera positiva sobre la competitividad, llegando a un equilibrio marginalmente distinto al *Equilibrio de Nash* tanto en el entorno de entrenamiento como en la de prueba. Este resultado va en línea con lo esperado y tiene un sustento económico subyacente, apoyando la teoría de que un mayor número de agentes desfavorece la coordinación entre los agentes de mercado e incrementa la competitividad, acercando de esta manera el equilibrio de mercado al *Equilibrio de Nash*.

Por último, una alternativa para resguardar los niveles de competitividad en los mercados cuando participan agentes autónomos es la inclusión de un agente altruista. Tal como se explica en la sección 3.2, el objetivo de este agente es fijar un precio lo más cercano posible al *Equilibrio de Nash* y de esta manera reducir el nivel de precios y rentabilidades obtenidas por los agentes. Para efectos de esta investigación, el agente altruista es diseñado para fijar un precio equivalente al *Equilibrio de Nash* en todo período t. Los resultados de este experimento muestran como el nivel de rentabilidad de los agentes se reduce de manera efectiva, llegando a niveles similares a los obtenidos con un mayor número de agentes tanto en el ambiente de entrenamiento como en prueba. Este resultado es interesante y va en línea con lo encontrado por [26], pues avala que incluir un agente altruista es suficiente para aumentar la competitividad del mercado, siendo esto válido incluso para escenarios donde se incluye variación en los costos de producción. Si bien este enfoque puede parecer simplista, incluir un agente altruista tiene dos posibles desventajas. Primero, se debe incurrir en el costo monetario de generar y mantener un agente que fije precios con bajos márgenes de rentabilidad, lo que podría ser de especial dificultad si no se dispone de recursos necesarios para asegurar la sostenibilidad de este sistema. Segundo, se debe diseñar un método que genere una predicción para el precio de Nash cada vez que ocurra un cambios en el nivel de precios. De esta forma, es esencial que los hacedores de política pública puedan abordar de manera oportuna ambas problemáticas para asegurar el correcto funcionamiento de esta solución.

Tabla 4.16: Resumen de resultados en configuración de entrenamiento.

Configuración	**N**	Δ				∇			
		μ	σ	d	Efecto	μ	σ	d	Efecto
$\rho = 0.003$	50	0.7258	0.0407	5.2011	Enorme	0.3388	0.0059	0.5967	Medio
$\rho = 0.002$	50	0.6912	0.0471	3.7347	Enorme	0.3386	0.004	0.5859	Medio
$\gamma = 0.7$	50	0.6128	0.0458	0.6150	Medio	0.3394	0.0023	0.6112	Medio
$\gamma = 0.8$	50	0.6116	0.0448	0.5805	Medio	0.3392	0.0027	0.6119	Medio
$k = 25$	50	0.6106	0.0453	0.5323	Medio	0.337	0.017	0.5335	Medio
$lr = 0.02$	50	0.6022	0.0464	0.1845	Muy pequeño	0.3202	0.0014	0.5969	Medio
Exp. Base	50	0.5976	0.052	-	-	0.3156	0.0548	-	-
$lr = 0.03$	50	0.5868	0.0479	0.4463	Pequeño	0.294	0.0063	0.1175	Muy pequeño
$k = 10$	50	0.5862	0.1164	0.1492	Muy pequeño	0.2964	0.1871	0.1467	Muy pequeño
$N = 3$	50	0.3132	0.044	11.5487	Enorme	0.0506	0.0042	6.8551	Enorme
$N = 5$	50	0.29986	0.0301	11.8446	Enorme	0.1300	0.0000	4.7916	Enorme
Sin inflación	50	0.2746	0.1664	2.5042	Enorme	0.2746	0.1664	0.3272	Pequeño
Agente altruista	50	0.0834	0.0595	11.0864	Enorme	-	-	-	-

Tabla 4.17: Resumen de resultados en configuración de prueba.

Configuración	**N**	Δ				∇			
		μ	σ	d	Efecto	μ	σ	d	Efecto
$\rho = 0.003$	50	0.4174	0.0801	0.7154	Medio	0.3228	0.0851	0.0390	Muy pequeño
$\rho = 0.002$	50	0.3944	0.0799	0.3753	Pequeño	0.3222	0.0880	0.0458	Muy pequeño
$\gamma = 0.8$	50	0.3814	0.0213	0.3124	Pequeño	0.3396	0.0019	0.3047	Pequeño
$\gamma = 0.7$	50	0.3734	0.0593	0.0933	Muy pequeño	0.3312	0.0607	0.0851	Muy pequeño
$lr = 0.02$	50	0.3718	0.0867	0.0527	Muy pequeño	0.3296	0.0928	0.0470	Muy pequeño
Exp. Base	50	0.3678	0.0671	-	-	0.3258	0.0639	-	-
$k = 25$	50	0.3550	0.0913	0.1574	Muy pequeño	0.3118	0.1012	0.1609	Muy pequeño
$lr = 0.03$	50	0.3114	0.1729	0.4857	Pequeño	0.2670	0.1712	0.5862	Medio
Sin inflación	50	0.2920	0.1405	0.6679	Medio	0.2920	0.1405	0.3047	Pequeño
$N = 5$	50	0.1594	0.0284	3.9890	Enorme	0.1422	0.0303	3.5227	Enorme
$k = 10$	50	0.1320	1.0029	0.3326	Pequeño	0.0818	1.0306	0.3330	Pequeño
$N = 3$	50	0.0900	0.0200	6.1856	Enorme	0.0520	0.0127	5.9809	Enorme
Agente altruista	50	-0.2332	0.0560	11.0778	Enorme	-	-	-	-

Capítulo 5

Conclusiones y Trabajo Futuro

5.1. Conclusiones

En un contexto social donde el uso de algoritmos de inteligencia artificial para las decisiones críticas está en auge, es natural preguntarse si esto podría tener consecuencias negativas a la sociedad. Un caso específico de esto son los algoritmos basados en Aprendizaje Reforzado y su capacidad para adaptarse a condiciones cambiantes. Desde esta vereda, esta tesis tiene por objetivo validar la hipótesis de que cuando estos algoritmos son usados para decisiones de *pricing*, estos capaces de mantener equilibrios no competitivos aun en escenarios con costos de producción variables.

Los resultados obtenidos validan parcialmente la hipótesis de investigación. La inclusión de la inflación en la configuración experimental muestra un impacto notable en las rentabilidades obtenidas por los agentes. Tanto en el entorno de entrenamiento como en el de prueba, los escenarios con inflación generaron mayores rentabilidades en comparación con aquellos sin inflación. Este hallazgo sugiere que la competitividad en los mercados puede verse afectada negativamente cuando los agentes económicos operan en un entorno inflacionario. Sin embargo, se observa que el grueso del impacto en la competitividad no se debe a un cambio en las estrategias de los agentes, sino al anclaje de los precios en los costos.

Por otro lado, los resultados no proporcionan evidencia suficiente para demostrar que los agentes aprenden estrategias de castigo efectivas para evitar el desvío del equilibrio. Solo en dos de los cincuenta experimentos ejecutados, los agentes respondieron adecuadamente al desvío del equilibrio. Este resultado contrasta con estudios previos que muestran una respuesta más robusta a los desvíos. Una posible explicación radica en las diferencias entre los algoritmos utilizados: el presente estudio se basa en DQN, mientras que estudios anteriores utilizan Q-learning, lo que podría explicar la variabilidad en las estrategias aprendidas.

Asimismo, hiperparámetros de los algoritmos como la tasa de aprendizaje y la tasa de descuento intertemporal tienen efectos diversos sobre la competitividad del mercado. Por ejemplo, un aumento en la tasa de aprendizaje puede incrementar la variabilidad de las acciones de los agentes, disminuyendo la estabilidad de los resultados. De igual manera, cambios en la configuración del modelo económico generan un impacto heterogéneo en los resultados. Un caso remarcable es el impacto del número de agentes en competencia, donde un mayor número de agentes impacta de forma negativa en las rentabilidades, acercándose a niveles

más competitivos. Estos resultados subrayan la importancia de contar con una configuración adecuada de hiperparámetros en los algoritmos de aprendizaje reforzado para mantener la estabilidad y competitividad del mercado.

Además, la inclusión de un agente altruista que fija precios cercanos al equilibrio de Nash demuestra ser una estrategia efectiva para reducir las rentabilidades de los agentes y aumentar la competitividad del mercado. Este hallazgo es consistente tanto en el entorno de entrenamiento como en el de prueba. Sin embargo, se identifican dos posibles desventajas de esta estrategia: el costo de generar y mantener el agente altruista, y la necesidad de predecir adecuadamente el precio de Nash ante variaciones en los niveles de precios.

Finalmente, los resultados de esta tesis sugieren la necesidad de continuar investigando el impacto de la variabilidad en los costos de producción en la estructura competitiva de los mercados. Además, se recomienda evaluar de manera empírica el alcance de las estrategias de castigo en función del algoritmo empleado. Por último, se destaca la importancia de diseñar políticas y regulaciones efectivas para asegurar la equidad y competitividad en mercados cada vez más influenciados por la inteligencia artificial y en entornos económicos dinámicos.

5.2. Trabajo Futuro

La presente tesis ha revelado varios hallazgos significativos sobre el impacto de la inflación y la competitividad en mercados influenciados por algoritmos de aprendizaje reforzado. No obstante, queda una vasta área de investigación por explorar. A continuación, se proponen varias direcciones para futuros trabajos:

- **Extensión del Modelo de Costos:** Un aspecto fundamental a investigar es la inclusión de otros tipos de variabilidad en los costos, como fluctuaciones en los precios de materias primas y cambios en la eficiencia de producción. Estos factores pueden proporcionar una visión más completa de cómo los agentes ajustan sus estrategias en diferentes contextos económicos.
- **Análisis de Políticas de Regulación:** Dado que los resultados sugieren que la competitividad del mercado puede verse afectada negativamente por la inflación, es crucial desarrollar y evaluar políticas regulatorias que mitiguen estos efectos. Estudios futuros podrían enfocarse en la implementación de regulaciones que promuevan una mayor transparencia y equidad en la fijación de precios algorítmica.
- **Implementación de Algoritmos Alternativos:** La investigación podría ampliarse mediante la comparación de diferentes algoritmos de aprendizaje reforzado, como el Proximal Policy Optimization (PPO) y el Soft Actor-Critic (SAC). Evaluar cómo estos algoritmos responden a la variabilidad en los costos y a los shocks inflacionarios puede aportar valiosos conocimientos sobre la robustez de diferentes enfoques de aprendizaje.

Estas propuestas de trabajo futuro no solo buscan ampliar el entendimiento teórico y práctico del pricing algorítmico en contextos económicos dinámicos, sino también contribuir al desarrollo de políticas y herramientas que aseguren mercados más justos y eficientes.

Bibliografía

[1] Tudor, C., "Integrated framework to assess the extent of the pandemic impact on the size and structure of the e-commerce retail sales sector and forecast retail trade e-commerce," Electronics, vol. 11, no. 19, p. 3194, 2022.

[2] Calvano, E., Calzolari, G., Denicolo, V., y Pastorello, S., "Artificial intelligence, algorithmic pricing, and collusion," American Economic Review, vol. 110, no. 10, pp. 3267–97, 2020.

[3] Klein, T., "Autonomous algorithmic collusion: Q-learning under sequential pricing," The RAND Journal of Economics, vol. 52, no. 3, pp. 538–558, 2021.

[4] Bertrand, J., "Review of "theorie mathematique de la richesse sociale" and of "recherches sur les principles mathematiques de la theorie des richesses."," Journal de savants, vol. 67, p. 499, 1883.

[5] Wikipedia, "Bertrand competition," 2005.

[6] Singh, N. y Vives, X., "Price and quantity competition in a differentiated duopoly," The Rand journal of economics, pp. 546–554, 1984.

[7] Amir, R., Erickson, P., y Jin, J., "On the microeconomic foundations of linear demand for differentiated products," Journal of Economic Theory, vol. 169, pp. 641–665, 2017.

[8] Dugar, S. y Mitra, A., "Bertrand competition with asymmetric marginal costs," Economic Inquiry, vol. 54, no. 3, pp. 1631–1647, 2016.

[9] Dubé, J.-P., "Microeconometric models of consumer demand," en Handbook of the Economics of Marketing, vol. 1, pp. 1–68, Elsevier, 2019.

[10] Davis, D. D. y Wilson, B. J., "Differentiated product competition and the antitrust logit model: an experimental analysis," Journal of Economic Behavior & Organization, vol. 57, no. 1, pp. 89–113, 2005.

[11] Duersch, P. y Eife, T. A., "Price competition in an inflationary environment," Journal of Monetary Economics, vol. 104, pp. 48–66, 2019.

[12] Varian, H. R., Intermediate microeconomics with calculus: a modern approach. WW norton & company, 2014.

[13] Sutton, R. S. y Barto, A. G., Reinforcement learning: An introduction. MIT press, 2018.

[14] Khamidehi, B. y Sousa, E. S., "Reinforcement learning-based trajectory design for the aerial base stations," en 2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), pp. 1–6, IEEE, 2019.

[15] Watkins, C. J. y Dayan, P., "Q-learning," Machine learning, vol. 8, pp. 279–292, 1992.

[16] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., *et al.*, "Human-level control through deep reinforcement learning," nature, vol. 518, no. 7540, pp. 529–533, 2015.

[17] Williams, R. J., "Simple statistical gradient-following algorithms for connectionist reinforcement learning," Machine learning, vol. 8, pp. 229–256, 1992.

[18] Schulman, J., Levine, S., Abbeel, P., Jordan, M., y Moritz, P., "Trust region policy optimization," en International conference on machine learning, pp. 1889–1897, PMLR, 2015.

[19] Nowé, A., Vrancx, P., y De Hauwere, Y.-M., "Game theory and multi-agent reinforcement learning," Reinforcement Learning: State-of-the-Art, pp. 441–470, 2012.

[20] Hernandez-Leal, P., Kartal, B., y Taylor, M. E., "A survey and critique of multiagent deep reinforcement learning," Autonomous Agents and Multi-Agent Systems, vol. 33, no. 6, pp. 750–797, 2019.

[21] Matignon, L., Laurent, G. J., y Le Fort-Piat, N., "Independent reinforcement learners in cooperative markov games: a survey regarding coordination problems," The Knowledge Engineering Review, vol. 27, no. 1, pp. 1–31, 2012.

[22] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., Silver, D., y Kavukcuoglu, K., "Asynchronous methods for deep reinforcement learning," en International conference on machine learning, pp. 1928–1937, PMLR, 2016.

[23] Schaul, T., Quan, J., Antonoglou, I., y Silver, D., "Prioritized experience replay," arXiv preprint arXiv:1511.05952, 2015.

[24] Van Hasselt, H., Guez, A., y Silver, D., "Deep reinforcement learning with double q-learning," en Proceedings of the AAAI conference on artificial intelligence, vol. 30, 2016.

[25] Calvano, E., Calzolari, G., Denicoló, V., y Pastorello, S., "Algorithmic collusion with imperfect monitoring," International Journal of Industrial Organization, vol. 79, p. 102712, 2021.

[26] Lepore, N., AI Pricing Collusion: Multi-Agent Reinforcement Learning Algorithms in Bertrand Competition. PhD thesis, Harvard University, 2021.

[27] Eschenbaum, N., Mellgren, F., y Zahn, P., "Robust algorithmic collusion," arXiv preprint arXiv:2201.00345, 2022.

[28] Abada, I. y Lambin, X., "Artificial intelligence: Can seemingly collusive outcomes be avoided?," Available at SSRN 3559308, 2020.

[29] Simon, H. A., "Altruism and economics," The American Economic Review, vol. 83, no. 2, pp. 156–161, 1993.

[30] Klimecki, O. M., Mayer, S. V., Jusyte, A., Scheeff, J., y Schönenberg, M., "Empathy promotes altruistic behavior in economic interactions," Scientific reports, vol. 6, no. 1, p. 31961, 2016.

[31] Hu, Y.-A. y Liu, D.-Y., "Altruism versus egoism in human behavior of mixed motives: An experimental study," American Journal of Economics and Sociology, vol. 62, no. 4, pp. 677–705, 2003.

[32] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., y

Riedmiller, M., "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.

[33] Kingma, D. P. y Ba, J., "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

[34] Cohen, J., Statistical power analysis for the behavioral sciences. Routledge, 2013.

[35] Benabou, R. y Gertner, R., "Search with learning from prices: does increased inflationary uncertainty lead to higher markups?," The Review of Economic Studies, vol. 60, no. 1, pp. 69–93, 1993.

[36] Fuess Jr, S. M. y Loewenstein, M. A., "On strategic cost increases in a duopoly," International Journal of Industrial Organization, vol. 9, no. 3, pp. 389–395, 1991.

[37] Tirole, J., The theory of industrial organization. MIT press, 1988.

[38] Nash, J., "Non-cooperative games," Annals of mathematics, pp. 286–295, 1951.

[39] Porter, M. E. y Strategy, C., "Techniques for analyzing industries and competitors," Competitive Strategy. New York: Free, 1980.

[40] Stiglitz, J., Greenwald, B., y Greenwald, B. C., Towards a new paradigm in monetary economics. Cambridge university press, 2003.

[41] Dong, X., Shen, J., Wang, W., Shao, L., Ling, H., y Porikli, F., "Dynamical hyperparameter optimization via deep reinforcement learning in tracking," IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 5, pp. 1515–1529, 2019.

Anexo A

Marco Teórico

Tabla A.1: Caracterización de series inflacionarias mensuales utilizadas para la ejecución de experimentos (%)

País	**Promedio**	**Std**	**Min**	**25 %**	**50 %**	**75 %**	**Max**
Canadá	0.16	0.38	-0.98	-0.10	0.13	0.39	1.15
China	0.19	0.62	-1.39	-0.20	0.10	0.54	2.60
Francia	0.12	0.32	-1.00	-0.10	0.11	0.33	1.08
Alemania	0.12	0.35	-1.01	-0.10	0.11	0.37	1.06
Italia	0.14	0.21	-0.58	0.00	0.19	0.27	0.62
Países Bajos	0.16	0.48	-1.00	-0.24	0.11	0.49	1.21
Singapur	0.13	0.50	-1.61	-0.20	0.10	0.44	2.01
Suecia	0.11	0.41	-1.34	-0.11	0.11	0.40	1.04
Suiza	0.04	0.35	-1.03	-0.11	0.00	0.20	1.12
Estados Unidos	0.18	0.38	-1.97	0.00	0.19	0.41	1.21

Anexo B

Metodología Experimental

B.1. Demostración ∇

Una de las debilidades de la ecuación expuesta en 3.23 es que no logra controlar por la diferencia en las tasas de crecimiento de la inflación y los equilibrios económicos (Nash y Monopolio). Para solucionar lo anterior, es posible imponer que el crecimiento de los equilibrios económicos sea igual a la tasa de inflación, es decir:

$$p_{t+1}^{N} = p_t^{N} \cdot (1 + \pi) \tag{B.1}$$

$$p_{t+1}^{M} = p_t^{M} \cdot (1 + \pi) \tag{B.2}$$

Si bien la solución anterior es capaz de controlar el efecto inflacionario sobre Δ_t, tiene la desventaja de no medirse contra los equilibrios teóricos. Esto es perjudicial para el análisis, pues impide medir las recompensas de los agentes frente al verdadero máximo global (equilibrio de monopolio).

Si denotamos por $(p_t^{N_F}, p_t^{M_F})$ a los precios de Nash y Monopolio forzados por las ecuaciones B.1 y B.2, podemos escribir la ecuación 3.23 como:

$$\nabla_t = \frac{\bar{R}_t - R_t^{N_F}}{R_t^{M_F} - R_t^{N_F}} \tag{B.3}$$

donde $(R_t^{N_F}, R_t^{M_F})$ corresponden a las rentabilidades obtenidas por los precios $(p_t^{N_F}, p_t^{M_F})$, respectivamente.

Sin embargo, los precios de Nash y Monopolio teóricos provienen de las ecuaciones 2.9 y 2.10. Si denotamos ambas soluciones teóricas por $(p_t^{N_T}, p_t^{M_T})$, podemos escribir la ecuación 3.23 como:

$$\Delta_t = \frac{\bar{R}_t - R_t^{N_T}}{R_t^{M_T} - R_t^{N_T}} \tag{B.4}$$

Adicionalmente, se genera una relación entre las rentabilidades $(R_t^{N_T}, R_t^{N_F})$ por medio de la siguiente expresión:

$$R_t^{N_T} = R_t^{N_F} + R_t^{N_\pi} \tag{B.5}$$

donde $R_t^{N_\pi}$ denota la fracción de la rentabilidad generada a partir de las diferencias entre la tasa de inflación y el crecimiento de los equilibrios de Nash y Monopolio.

A partir de las ecuaciones anteriores, es posible reescribir Δ_t^T:

$$\begin{aligned}
\Delta_t^T \cdot \frac{R_t^{M_T} - R_t^{N_T}}{R_t^{M_F} - R_t^{N_F}} &= \frac{\bar{R}_t - R_t^{N_T}}{R_t^{M_T} - R_t^{N_T}} \cdot \frac{R_t^{M_T} - R_t^{N_T}}{R_t^{M_F} - R_t^{N_F}} \\
&= \frac{\bar{R} - (R_t^{N_F} + R_t^{N_\pi})}{R_t^{F_M} - R_t^{N_F}} \\
&= \nabla_t - \frac{R_t^{N_\pi}}{R_t^{M_F} - R_t^{N_F}} \\
\Delta_t &= (\nabla_t - \text{IE}) \cdot \beta_t
\end{aligned} \tag{B.6}$$

donde Δ_t mide el nivel de competitividad de los mercados usando los valores teóricos de Nash y Monopolio, ∇_t captura las ganancias en Δ_t controlando por la disparidad en los crecimientos (en otras palabras, captura sólo ganancias generadas por estrategias cooperativas), IE indica las ganancias en Δ_t por el efecto inflación, y β_t puede ser interpretado como el ratio entre los equilibrios teóricos y forzados.