Distribution Agreement

Shuqian Ni Apr 9, 2025

Algorithmic Self-Preferencing in E-Commerce: Analyzing Amazon's Buy Box Allocation Across Different Seller Types and its Antitrust Implications

by

Shuqian Ni

Maria Arbatskaya

Adviser

Economics

Maria Arbatskaya

Adviser

Hau Nguyen

Committee Member

Keith Robinson

Committee Member

2025

Algorithmic Self-Preferencing in E-Commerce: Analyzing Amazon's Buy Box Allocation Across Different Seller Types and its Antitrust Implications

by

Shuqian Ni

Maria Arbatskaya

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2025

Abstract

Algorithmic Self-Preferencing in E-Commerce: Analyzing Amazon's Buy Box Allocation Across Different Seller Types and its Antitrust Implications

By Shuqian Ni

This thesis investigates algorithmic self-preferencing on Amazon by analyzing Buy Box allocation across different seller types. Key contributions include introducing referral fees and offer count as novel explanatory variables to capture Amazon's commission-based incentives and platform competition intensity, and conducting a cross-country comparison with an underexplored market—Japan. Using product-level data from Keepa across the U.S., France, and Japan, we used logistic regression models with bootstrapped AME to evaluate how referral fees, pricing, market structure, and quality metrics influence Amazon's probability of winning the Buy Box. Robustness is confirmed through LASSO regression and 4 alternative model specifications.

Findings show Amazon is more likely to win the Buy Box in high-referral-fee categories, suggesting a strategic incentive to prioritize market dominance over short-run commission revenue. Secondly, the negative association between the current Buy Box price and Amazon's probability of winning Buy Box is weaker in high-referral-fee categories, revealing an internal trade-off between commission revenue and price margin. Thirdly, in low-referral-fee categories, Amazon is more likely to win the Buy Box without offering the lowest price over the past 90 days. This indicates stronger algorithmic favoritism in low-referral-fee categories, where the platform's algorithm disproportionately favors its own retail offer in Buy Box allocation even when third-party sellers offer more competitive prices. Overall, this thesis offers new insights into the economic and strategic motivations behind algorithmic self-preferencing and underscores the need for more tailored, tier-sensitive antitrust regulatory responses.

Algorithmic Self-Preferencing in E-Commerce: Analyzing Amazon's Buy Box Allocation Across Different Seller Types and its Antitrust Implications

By

Shuqian Ni

Maria Arbatskaya

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Economics

2025

Acknowledgements

I would like to begin by expressing my sincere gratitude to all of my committee members who have guided me throughout my college and the thesis process. From the classroom of Econ 220 to the founding of HOX, and now as a member of my committee, Dr. Hau Nguyen has always encouraged me to believe in myself and achieve more. Dr. Keith Robinson's exceptional expertise in antitrust law and policy design has elevated this thesis to a deeper, more impactful level, showing me how economics can shape meaningful change in the world. My deepest gratitude goes to my adviser, Dr. Maria Arbatskaya, who sparked my passion for industrial organizations, and gave me an incredible amount of trust and support along the way; it is truly an honor to be her student and complete my thesis under her guidance.

I also wish to thank Wonjun Choi and Alessia Scudiero for closely following my thesis progress and for introducing me to new analytical tools and expressions that helped sharpen my findings and insights. I could not have completed this thesis as smoothly without your help.

My warmest gratitude goes to my family. Their love and care over the past 22 years have been my biggest asset and the greatest source of strength in facing challenges and turning them into milestones.

Finally, I want to thank my friends who offered their encouragement, patience, and support for my thesis process. I am especially thankful to Monica Cai, Kiki Gu, Peter Li, and Rita Xu for being my companions during long writing sessions. I also want to extend my thanks to Helen Jin, Minsol Kim, Daniel Li, and Tobby Yang for listening to my defense and always being there for me. I am incredibly lucky to have had all of you throughout this journey.

Contents

Algorithmic Self-Preferencing in E-Commerce: Analyzing Amazon's Buy Box Allocation Across Different Seller Types and its Antitrust Implications

Shuqian Ni

April 2025

1 Introduction

1.1 Motivation

The rapid growth of e-commerce has fundamentally reshaped the global retail landscape, transforming the way consumers search for and purchase products. Among the many e-commerce platforms, Amazon has emerged as a dominant player. In key markets such as the United States, Amazon captured 37.6% of the e-commerce market in 2023, far surpassing competitors like Walmart, which held just 6.4% market share (Statista, 2024). Globally, it has the largest e-commerce net sales of 152.84 billion U.S. dollar in 2024 (Statista, 2025). Amazon's dominance, however, raises concerns regarding fair competition and the role that Amazon's algorithmic systems play in shaping market outcomes.

Unlike traditional retailers, Amazon functions as both a platform operator and a direct market participant, integrating offerings from independent third-party sellers and its own first-party products. This dual role creates an inherent conflict of interest, as Amazon holds considerable control over platform rules (Khan, 2017). As a result, concerns over algorithmic self-preferencing, where Amazon systematically favors its own offerings, have received substantial research attention, such as a search ranking algorithm (Farronato et al., 2023), a pricing algorithm (Rory & Aggarwal, 2023), and a recommendation algorithm (N. Chen & Tsai, 2019). While the use of algorithms is widely recognized, their impact on competition and market dynamics continues to be a pressing concern.

1.2 Buy Box and its Anti-trust Concerns

In this thesis, we focus on one of Amazon's most influential algorithms—the Buy Box Algorithm. The Buy Box, often referred to as the "Featured Offer," is the display from which shoppers can directly "Add to Cart" or "Buy Now" a selected product offering, highlighted in Figure 1's red box. (Federal Trade Commission, 2023)

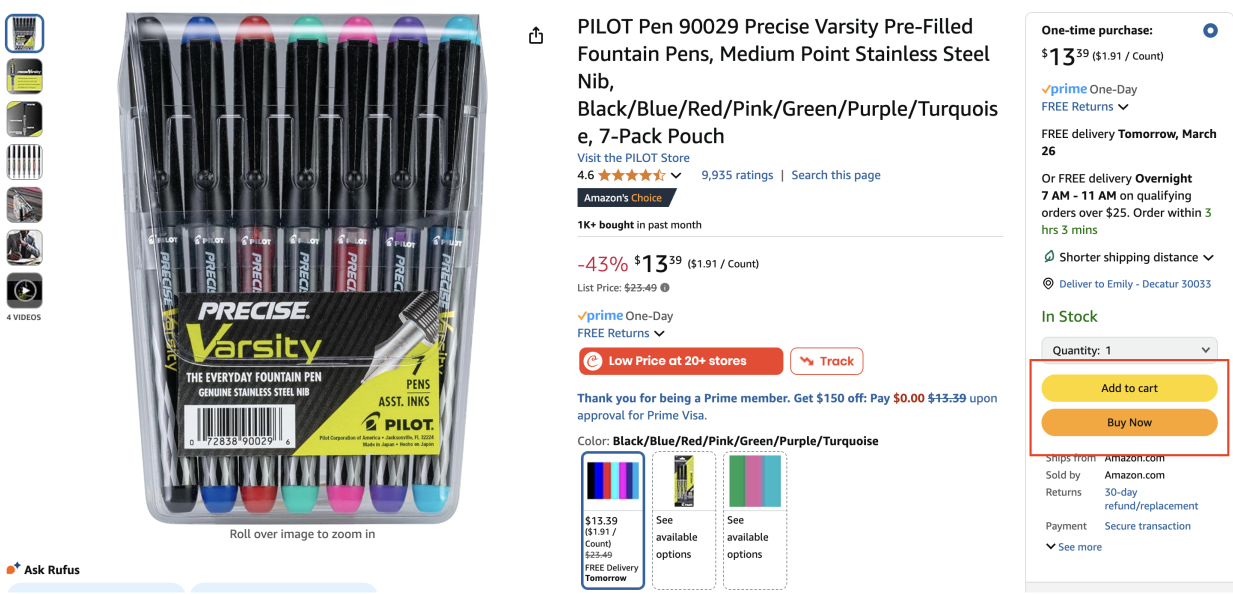


Figure 1: Item with Buy Box (Desktop)

When more than one seller lists the same item, only one wins the Buy Box, while others remain accessible only through scrolling down the buy box, and taking additional clicks on the detailed product page.

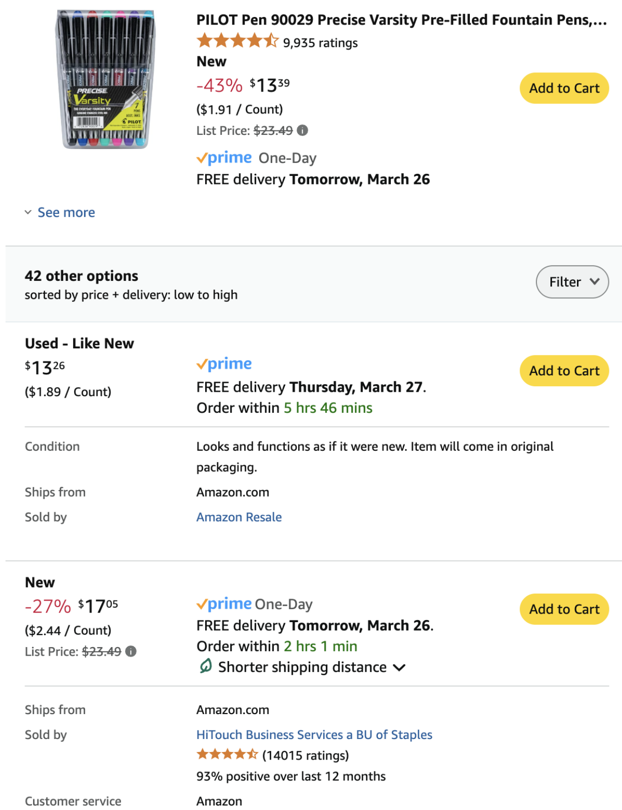


Figure 2: Item without Buy Box (Desktop)

Given that more than 80% of Amazon's sales are made through the Buy Box (C. E. Etumnu & Malone, 2024; Federal Trade Commission, 2023), it becomes the determining factor of whether sellers can access consumers on Amazon or not, making it crucial for sellers to secure the Buy Box position. If Buy Box allocation is not a result of fair competition, it will raise serious anti-trust concerns. Firstly, Amazon's dual role as both a platform operator and a first-party seller creates incentives for Amazon to self-preference by disproportionately favoring Amazon or Amazon's preferred partners to be the buy box winner. In situations where such self-preference actions promote inferior products or make it difficult for consumers to find their preferred options, unfair Buy Box allocations could potentially harm consumers by reducing consumer welfare or limiting transparency (Farronato et al., 2023; Padilla et al., 2022). Secondly, Amazon's Fulfillment by Amazon (FBA) program plays a crucial role in Buy Box allocation (Amazon, n.d.; Raval, 2023), raising concerns about tying practices. To secure Buy Box ownership, third-party sellers are pressured to sign up for FBA services to stay competitive, allowing Amazon to condition sellers' access to the platform's marketplace (Khan, 2017). Thirdly, unfair Buy Box allocation can also distort broader e-commerce competition by restricting sellers' ability to participate on rival platforms (Federal Trade Commission, 2023). Amazon enforces strict price monitoring to ensure that sellers offer the lowest price on Amazon compared to other e-commerce platforms. If a seller lists a lower price on a competing platform, Amazon removes its Buy Box position, significantly reducing its visibility and sales potential. This creates a strong disincentive for sellers to price competitively on other platforms, making it harder for rival e-commerce marketplaces to attract sellers and compete fairly. Research has shown that Amazon's price parity policies influence seller behavior across multiple platforms, effectively suppressing price competition and reinforcing Amazon's dominance (Hunold et al., 2022).

Regulatory authorities have taken notice of these concerns. Beginning in 2019, The European Commission (EC) started an investigation into Amazon's use of non-public data from third-party sellers and its biased criteria for Buy Box and Prime. By 2022, the EC concluded that Amazon grants preferential treatment to its own retail offerings and sellers use FBA, which disturbs competition and abuses market dominance (European Commission, 2022). In response, Amazon promised not to use marketplace seller data and to ensure non-discriminatory access to Buy Box and Prime. However, these commitments are only applied within the European Economic Area (EEA), and concerns about Buy Box favoritism continue in other geographic areas.

Following the EC's findings, the U.S. Federal Trade Commission (FTC) started an antitrust complaint against Amazon in 2023 for manipulating the Buy Box algorithm to favor its own products, pressuring third-party sellers to use FBA services, and suppressing cross-platform price competitions. If these allegations are proven, Amazon's conduct would violate Section 5(a) of the FTC Act, which prohibits unfair methods of competition, and Section 2 of the Sherman Act, which forbids monopoly maintenance through exclusionary conduct (Federal Trade Commission, 2023).

More recently, the Japan Fair Trade Commission (JFTC) also launched an investigation against Amazon Japan's Buy Box. Similar to the European and United States, JFTC alleges Amazon for pressuring sellers to offer the lowest price at Amazon and to adopt FBA, which potentially violating Japan's Antimonopoly Act (The Japan Times, 2024). While the case remains ongoing, the JFTC's continued scrutiny further confirms abuses of the Buy Box algorithm and its severity in different geographic locations.

In summary, the actions by the EC, FTC, and JFTC illustrate a growing consensus among global regulators that Amazon's Buy Box allocation is not the outcome of fair competition, highlighting the urgent need for empirical analysis of Amazon's Buy Box practices and their antitrust impact.

1.3 Research Objectives

This thesis aims to contribute to this urgent need by examining Amazon's Buy Box allocation across different seller types and offering new insights into Amazon's algorithmic self-preferencing. As defined by the Global Competition Review (Duquesne et al., 2002) and supported by Khan (2017), Farronato et al. (2023), and Raval (2023), algorithmic self-preferencing refers to a platform's use of algorithms to systematically favor its own products or services over those of third-party sellers. In this thesis, we define self-preferencing as a disproportionately high probability of Amazon Retail winning the Buy Box after controlling for observable product and market characteristics. While this thesis does not make causal claims about the presence or mechanisms of algorithmic self-preferencing, it identifies patterns that align with such behavior, providing a framework for evaluating whether observed Buy Box outcomes reflect fair competition or are shaped by Amazon's underlying profit-maximizing incentives. Specifically, we aim to answer two key questions:

1. Does Amazon's probability of winning the Buy Box vary with low/high referral fee levels in the Grocery Category?

2. How do factors associated with Amazon's Buy Box winning probability vary across different countries in different continents (France, Japan, and the United States)?

The category selection is informed by C. E. Etumnu and Malone (2024)'s empirical analysis of Buy Box allocation in the grocery sector. Building on their work, we re-examine Buy Box outcomes using updated data in this fast-moving consumer goods category, with a particular focus on incorporating Amazon's commission revenue incentives through referral fee structures. Our study expands the geographic scope by incorporating Japan, which is a market that has received limited attention in cross-country analyses, to better understand how Buy Box dynamics vary across regions with distinct regulatory and competitive environments. In addition, we implement logit regression models with a set of other empirical approaches to highlight differences in Buy Box allocation between Amazon and non-Amazon sellers (FBA and FBM), enhancing the interpretability of seller-type effects.

2 Literature Review

As Amazon's Buy Box directly determines which product is prominently displayed, understanding the role of prominence and its impact on consumer search behavior provides important context for understanding the implications of Buy Box. Research in this area has consistently demonstrated that prominence significantly influences consumer behaviors and seller profits. Wolinsky (1986) revealed that prominent firms often charge lower prices but can still earn higher profits due to greater market share in the random search models, suggesting prominence's impacts on consumer purchasing behaviors. Expanding on this, under the homogeneous products assumption with different consumer search costs, Arbatskaya (2007) concludes that product prices often decline with lower-ranking positions, showing how search orders may influence pricing strategies in an online marketplace. Armstrong and Zhou (2011) further explored markets with search frictions and found that sellers benefit when consumers see their products first, as search costs lead shoppers to settle for satisfactory options rather than searching for the best deal.

In the online context, Y. Chen and He (2011) examined paid placement in search engines and found that sellers with better products (e.g., more relevant, better quality products) tend to invest in higher listing positions. When consumers trust a platform's ranking system and interpret the position as a signal of relevance, they might be more likely to click on prominent listings. Choi and Mela (2019) provided additional evidence by showing that higher-ranked and more visible products receive more browsing and clicks.

Given the power of prominence in impacting market dynamics, it is important to examine if Amazon conducts self-preferencing in the Buy Box algorithm. Existing research has provided valuable insights into this topic. The foundational work of Khan (2017) in Amazon's Antitrust Paradox set the stage for treating e-commerce self-preferencing as an antitrust issue, arguing that Amazon's dual role creates an inherent conflict of interest and strong incentives to favor its own offerings. Subsequent studies have expanded on this argument by examining how Amazon's algorithms foster self-preferencing. For instance, N. Chen and Tsai (2019) demonstrate that Amazon's recommendation system systematically promotes its own products in the "Frequently Bought Together" feature, increasing their visibility and sales. Similarly, Farronato et al. (2023) find that Amazon's search engine algorithm gives preferential treatment to its private-label brands by boosting their rankings and tailoring search results to direct consumers toward Amazon-owned products. Lill et al. (2024) further illustrate how the Badge algorithm amplifies self-preferencing by disproportionally favoring products sold by Amazon by representing them more often in Amazon Choice and Best Seller Badge, resulting in higher prices paid by shoppers.

While these algorithms operate independently, they collectively shape Amazon's marketplace by influencing product visibility, pricing dynamics, and consumer decision-making. Among them, the Buy Box algorithm plays a critical role, as it directly determines which seller's offer is displayed the most to shoppers (Federal Trade Commission, 2023). Unlike standalone pricing or ranking algorithms, the Buy Box integrates multiple factors into a complex decision-making process that significantly influences sales outcomes (Buchi, 2025). Given the Buy Box algorithm's importance and complexity, extensive research has identified the factors that determine Buy Box allocation, including price (Raval, 2023), fulfillment method (Jürgensmeier et al., 2024), algorithmic ranking biases (L. Chen et al., 2016), and regulation intervention (Rottembourg, 2024).

One of the key but relatively underexplored factors influencing self-preferencing behavior is the platform's profit-maximization incentives. Theoretical research suggests that while platforms have the ability to engage in self-preferencing, it is not always an optimal strategy, particularly when they can generate revenue through alternative means such as commissions and sponsored advertising (Long & Amaldoss, 2024). In this context, self-preferencing could be influenced by the platform's financial incentives. Despite the growing body of research on Amazon's Buy Box algorithm, existing literature has largely overlooked how Amazon's economic interests, particularly its ability to extract revenue from third-party sellers through referral fees, affect its incentives to self-preference by using Buy Box. This thesis addresses this gap by categorizing product markets into high and low referral fee groups to analyze whether Amazon's likelihood of winning the Buy Box differs across these categories.

Another gap in the current literature is the use of product-level data. Existing studies primarily rely on offer-level data, which captures individual seller listings rather than broader market dynamics at the product level. A notable exception is C. E. Etumnu and Malone (2024), who analyzed product-level data by adopting a discrete choice model to analyze the Buy Box seller assignment in the grocery sector. Their findings challenge the previous findings that Amazon's algorithm is purely anti-competitive, showing that being the lowest-priced seller over a 180-day period is the most significant factor in securing Buy Box ownership. They also find that low current prices and low out-of-stock rates play critical roles, confirming the complexity of Buy Box allocation beyond simple self-preferencing claims. However, their study does not incorporate competition metrics such as the number of competing offers, leaving an important gap in understanding how market competition influences Buy Box outcomes and Amazon's self-preferencing. This thesis builds on their work by deepening the analysis on the grocery category, incorporating competition intensity through the number of offers, and examining whether Amazon's probability of winning the Buy Box decreases as the number of third-party sellers increases.

In addition to refining the competitive framework of Buy Box research, this thesis also introduces a comparative analysis across three major markets: the United States, France, and Japan. Most prior studies have examined Amazon's self-preferencing within the U.S. or European contexts, but little research has focused on this issue within the Japanese market. The European Commission has taken a more interventionist approach to digital market regulation, with stricter oversight of platform favoritism (European Commission, 2022). France, as part of the European Union, has been subject to increased regulatory scrutiny, which may influence Amazon's approach to Buy Box allocation in this market (Rottembourg, 2024). In contrast, Japan's regulatory landscape has historically been less stringent, with the Japan Fair Trade Commission (JFTC) only launching an investigation into Amazon's Buy Box practices in November 2024 (The Japan Times, 2024). Given its relatively recent regulatory focus, Japan's Amazon market may exhibit stronger self-preferencing tendencies compared to earlier regulated markets. By comparing these three markets, this thesis seeks to determine whether Amazon's Buy Box algorithm and the degree of self-preferencing differ in different regulatory environments.

Overall, this thesis builds on the existing literature by shifting the focus from offer-level to product-level competition analysis, expanding the geographic scope to include Japan, and introducing competition metrics such as the number of competing offers and referral fees. These contributions provide a more comprehensive understanding of how Amazon's Buy Box algorithm interacts with market competition and regulatory environments, offering new insights into the conditions under which self-preferencing is more pronounced and how to propose more targeted policies.

3 Hypothesis

Building on existing literature, this thesis investigates the determinants of Buy Box seller allocation with a particular focus on the effects of competition intensity, pricing, quality metrics, inventory levels, current sales rank, referral fees, and geographic differences. Below, we outline key hypotheses based on prior empirical findings and economic theory.

3.1 Competition Intensity

We hypothesize that higher competition reduces the probability of Amazon retaining the Buy Box but increases the probability of third-party sellers winning the Buy Box. The rationale behind this hypothesis is that higher competition intensity provides customers with more options and increases the probability that customers prefer other sellers over Amazon. When a considerable number of customers prefer other sellers over Amazon for a specific product, to remain attractive as a platform operator, Amazon will be less incentivized to implement self-preferencing at the risk of losing platform users. Instead, presenting the most attractive seller and collecting commissions from it could be a more profit-optimal option. Therefore, higher competition might indirectly restrict Amazon's incentives and ability to self-preference, thus creating more opportunities for third-party sellers to be presented in front of shoppers.

In this thesis, the number of offers (`Offer_Count`) is the key indicator of current competition intensity. Following the previous rationale, we hypothesize that a higher number of competing offers suggests a higher current competition level for this product, therefore reducing Amazon's probability of becoming a Buy Box seller and increasing other types of sellers' probability of winning Buy Box.

3.2 Price

The price of a product plays a crucial role in determining Buy Box ownership (Amazon, n.d.). In this thesis, the current Buy Box Price `Buy_Box_Price` and the identity of the 90-day Lowest Price seller `90_days_Lowest_Price_Seller` are two separate indicators for measuring price.

The current Buy Box Price is the price of the Buy Box winner that is listed on the product page at the given moment, including shipping, packaging, and other fees, which reflects the total sales price currently. Following Raval (2023)'s findings that Amazon Retail receives significant preferential treatment in the Buy Box algorithm, with a 16% price advantage over FBA sellers and a 46% advantage over FBM sellers, we hypothesize a lower current Buy Box price increases the likelihood of a seller winning the Buy Box, with the effect possibly weaker for Amazon than for third-party sellers, as third-party sellers may need to compete more aggressively on price to win the Buy Box.

The `90_days_Lowest_Price_Seller` is the identity of the Buy Box winner that maintains the lowest price over the past 90 days. In contrast to the current Buy Box Price hypothesis, empirical evidence provided by C. E. Etumnu and Malone (2024) suggested that sellers typically win Amazon's Buy Box by offering the lowest price over the past 180 days, with no substantial self-preferencing effects. Following this result and our acknowledgment of Amazon's incentives to show the most attractive offer to remain competitive as a platform, we suggest if a seller consistently has the lowest price over 90 days, they could have a higher probability of winning the Buy Box, regardless of the seller's identity.

This distinction between short-run and long-run price metrics captures two different situations where pricing influences the Buy Box assignment. A possible explanation is that immediate price adjustments in the short run may allow self-preferencing, but the platform rewards sustained competitive pricing regardless of sellers' identities in order to remain attractive to shoppers.

3.3 Quality Metrics - Reviews

We examine the role of review ratings and the number of reviews as quality metrics that are used when assigning Buy Box.

Firstly, the review rating `Review_Rating` is used as the direct metric to assess product quality. We hypothesize that higher-rated products are more likely to win the Buy Box due to their stronger customer approval. Review ratings reflect customer satisfaction, which aligns with Amazon's goal of recommending the most desirable products to shoppers.

Secondly, the number of reviews `Review_Count` is the key indicator of seller credibility and engagement. We hypothesize that a higher number of reviews is positively correlated with the probability of getting Buy Box, as a larger volume of reviews serves as social proof of consumer trust (Cialdini, 1993), reducing uncertainty in purchasing decisions. Given that Amazon's Buy Box algorithm is designed to optimize customer experience, it is likely that sellers with higher review counts gain an advantage in Buy Box allocation. However, this effect may not be strictly linear. If a product has a high number of reviews but most of them are negative (with low review ratings), it could weaken its chances of securing the Buy Box. Therefore, we propose an alternative hypothesis that while review volume generally increases Buy Box probability, this effect may diminish or reverse if the product has consistently low ratings.

3.4 Inventory Levels

As Amazon's Buy Box Eligibility policy states an out-of-stock situation would typically result in losing Buy Box (Amazon, n.d.), we hypothesize that a higher out-of-stock rate decreases the probability of winning Buy Box. However, reverse causality concerns might arise, as frequently winning the Buy Box makes the product popular, and being selected as the Buy Box seller increases the likelihood of running out of stock due to high demand. To address this, we use the 90-day out-of-stock rate `90_days_OOS` instead of the current stock status to ensure our variable captures a longer-term pattern of inventory availability rather than short-term fluctuations. Following the previous rationale, we hypothesize that a higher out-of-stock rate over the past 90 days decreases the probability of winning the Buy Box.

3.5 90-days Sales Rank

The sales rank of a product on Amazon is collected in this thesis as an indicator of sales performance (C. Etumnu & Noumir, 2023). There are two types of sales rank in Keepa, one is the primary category, such as grocery; another is the subcategory, such as breakfast cereal. In this thesis, the sales rank is collected within its primary category: for example, a grocery product's sales rank indicates its position among other grocery items, while a book's sales rank reflects its ranking within the book category. We hypothesize that products with a lower (better) sales rank represent stronger sales performance and are more likely to win the Buy Box, as Amazon's Buy Box algorithm is designed to select products that benefit shoppers' experience. However, a key concern is reverse causality. A lower sales rank improves a seller's chances of winning the Buy Box, but at the same time, being the Buy Box winner promotes visibility and increases sales, which raises reverse causality concerns. To address this issue, we use the lagged sales rank (`90_Sales_Rank`) to monitor and correct for potential endogeneity.

3.6 Referral Fees

Referral fees work as commissions that third-party sellers pay to Amazon in order to list items on the platform. We hypothesize that Amazon's self-preferencing behavior differs based on referral fee structures. In high referral fee categories, Amazon may have higher incentives to collect more referral fees to earn more profits, potentially reducing self-preferencing behaviors. Therefore, we hypothesize that higher referral fees reduce the probability of Amazon winning the Buy Box but increase the probability of third-party sellers winning the Buy Box. Conversely, when referral fees are low, Amazon faces less commission revenue from third-party sellers and thus may have stronger incentives to promote its own retail listings, resulting in greater self-preference.

In Groceries, referral fees have a fixed threshold across three countries of interest. For example, in the United States, Grocery products with a total sales price of $15 or less pay a referral fee rate of 8%, but for Grocery products with a total sales price greater than $15, a 15% referral fee rate is charged. Accordingly, we classify Grocery data as a binary low/high referral fee variable(`Referral_Fee_Binary`), allowing analysis of how referral fees might affect Buy Box allocation.

3.7 Geographic Differences

We hypothesize that Amazon's self-preferencing behavior varies across different geographic areas. Amazon's self-preferencing might be stronger in geographic areas that have weaker regulations or enforced more recently. This thesis focuses on three countries: France, the United States, and Japan.

The European Commission (EC) is the first regulation organization that investigates Amazon for self-preferencing. EC's case started in 2019, and led to legally binding commitments in December 2022 (European Commission, 2022), requiring Amazon to treat all sellers equally in the Buy Box ranking within the European Economic Area (EEA). Therefore, we expect self-preferencing to be weakest in France, where these regulations are already enforced.

In the United States, regulatory scrutiny of Amazon's Buy Box practices has increased more recently with the Federal Trade Commission (FTC) filing an antitrust lawsuit in 2023, but no regulatory commitments have yet been implemented (Federal Trade Commission, 2023). Consequently, we expect self-preferencing in the U.S. to be stronger than in France but weaker than in Japan, where regulatory action has been more delayed.

In Japan, the Japan Fair Trade Commission (JFTC)'s regulation on Amazon has historically been weaker when compared to the other two areas. While Amazon Japan was investigated in 2017 for anti-competitive price parity clauses, it was not until November 2024 that the JFTC launched a direct investigation into Buy Box favoritism (The Japan Times, 2024). This suggests that Amazon Japan has likely engaged in stronger self-preferencing for a longer period, as regulatory intervention has only recently intensified.

Due to different levels of regulatory scrutiny and enforcement timelines for these three geographic areas, within this thesis' scope, we hypothesize that Amazon's self-preferencing in Buy Box seller allocation is the strongest in Japan, followed by the United States, and the least prominent in France.Building on existing literature, this thesis investigates the determinants of Buy Box seller allocation with a particular focus on the effects of competition intensity, pricing, quality metrics, inventory levels, current sales rank, referral fees, category differences, and geographic differences. Below, we outline key hypotheses based on prior empirical findings and economic theory.

4 Data and Variable Description

4.1 Data Source and Collection

The primary data source for this thesis is Keepa.com, a platform that tracks metrics for over 3 million Amazon products, and provides detailed product categorization across multiple countries. This thesis focuses on Grocery category and spans several geographic markets, including the United States, France, and Japan.

This study uses product-level data collected from the Amazon marketplaces in the United States (amazon.com), France (amazon.fr), and Japan (amazon.jp) between January 20 and February 28 in 2025. Product level refers to information captured at the individual product listing level, with each row in the dataset representing a unique product identified by its Amazon Standard Identification Number (ASIN). Using the Keepa API, I extracted up to 5,000 data of the top-ranked products for each subcategory within the Grocery category. To maintain consistency and reduce variation noise, the top 5,000 sales rank refers to the rank in the subcategory, and only one variation (i.e. size, color) per product was included.

4.2 Data Cleaning

To support cross-country analysis of Amazon's Buy Box allocation, we standardized product-level data from Keepa by merging subcategories into country-level datasets (e.g., *US Grocery Combined*). A structured cleaning pipeline was applied to retain only relevant variables such as seller type, price, review metrics, sales rank, and stock availability. Listings marked as "Buy Box Unqualified" or missing critical data were excluded to ensure analytical consistency. Specifically, 33.58% is filtered in the United States, 33.33% is filtered in France, 57.93% is filtered in Japan.

Binary variables were constructed to capture key theoretical constructs. The dependent variable, `Amazon_Binary`, identifies whether Amazon won the Buy Box. `BB_Winner_is_LP_Seller` captures price competitiveness, and `Referral_Fee_Binary` classifies products into high- or low-referral-fee tiers, based on country-specific thresholds (see Appendix A.1.2 for full details on threshold policy).

Categorical variables were also created to reflect market structure and supply reliability. `Offer_Count_Categorized` defines four levels of seller competition, adapted to each category's distribution while maintaining consistent bin counts across countries. `OOS_Categorized` captures five levels of 90-day out-of-stock rates, using the U.S. thresholds across all countries to ensure comparability.

Several continuous variables were transformed to address skewness and improve interpretability. Sales rank was log-transformed, and review counts were scaled in thousands (`Review_Count_Thousand`). Buy Box price was log-transformed and mean-centered to create `Centered_Log_Buy_Box_Price`, which helps mitigate multicollinearity with `Referral_Fee_Binary` and enhances interpretability of interaction effects.

Two interaction terms were included to test key theoretical mechanisms. The interaction between `Centered_Log_Buy_Box_Price` and `Referral_Fee_Binary` examines whether Amazon's self-preferencing behavior varies across referral fee tiers and price levels. Similarly, the interaction between `Avg_Review_Rating` and `Review_Count_Thousand` captures the credibility of social proof, based on the assumption that high review volume strengthens the signal of high ratings (Cialdini, 1993).

Importantly, typical outliers such as products with extremely high review counts or Buy Box prices were not removed, as these values reflect real market situations rather than data entry errors.

Together, these constructions lay the foundation for the regression analysis that follows. The full variable construction logic and cleaning steps are provided in Appendix A.1 for replication and transparency.

After data cleaning, the numbers of observations retained in the regression analysis are: 6040 in the United States, 1216 in France, and 7304 in Japan.

4.3 Interaction Variables

The interaction term, `Centered_Log_Buy_Box_Price` × `Referral_Fee_Binary`, is created to test a moderation effect on whether the effect of price on Amazon's likelihood of self-selling varies at different referral-fee levels. The inclusion of this interaction term is motivated by a profit-maximizing logic. Amazon, as a dual-role platform, receives profit from both commission and direct margins. This setup aligns with economic models of platform strategy under dual roles, such as those proposed in Tremblay and of Business (2021), where margin-based incentives are determined by both price and commission rates, which eventually guide platform allocation decisions.

Our hypothesis is that, in high-referral-fee categories, the negative effect of price on Amazon's Buy Box success will be weaker. The rationale is that, even if Amazon loses the Buy Box due to a higher price, it still earns a substantial commission from third-party sellers, thereby reducing the penalty for setting a higher price. As a result, this allows Amazon to have greater pricing flexibility, which means Amazon has more room to adjust its prices without significantly lowering its chances of winning the Buy Box. In contrast, in low-referral-fee categories, where Amazon earns relatively less from commissions, the incentive to win the Buy Box directly is stronger. Consequently, we expect the negative association between price and Buy Box probability to remain strong in these categories, aligning with typical price-competitiveness logic.

Another interaction term in this thesis is `Avg_Review_Rating` × `Review_Count_Thousand`. This term is created based on the hypothesis that the credibility of reviews depends on both their volume and their average rating. A product with a perfect 5-star rating but only 2 reviews may not signal quality as strongly as a product with a 4.6-star rating and over 3,000 reviews. Amazon may be more inclined to win the Buy Box for highly credible products that combine strong average ratings with large volumes of customer feedback.

4.4 Descriptive Statistics

Before proceeding to regression analysis, we first examine the summary statistics of key variables across the U.S., France, and Japan. Table 2 presents the count, mean, standard deviation, minimum, and maximum for each variable included in the main models. Detailed versions by country can be found in Tables 3, 4, and 5.

The binary dependent variable `Amazon_Binary` presents an interesting pattern. Amazon is more active as a Buy Box winner in France and United States: mean = 0.17 for France, mean = 0.15 for the United States, and mean = 0.06 for Japan. This is surprising because Japan's antitrust scrutiny of Amazon is relatively less active, but the descriptive summary table exhibits the lowest rate of Amazon Buy Box ownership. This pattern might suggests that category-level dynamics and local market situations may be more influential than regulatory differences in shaping Amazon's self-preferencing strategy.

The variable `Referral_Fee_Binary` is a core explanatory variable in this analysis. Its distribution reveals a consistent pattern across countries: a majority of grocery products fall into the high-referral-fee category. Specifically, the proportion of high-fee products is mean = 0.59 in the United States, mean = 0.63 in France, and mean = 0.25 in Japan. The skew toward high-fee categories in the U.S. and France suggests that Amazon's marketplace structure in Grocery relies heavily on categories where referral fees are at the higher tier (e.g., 15%). Even in Japan, despite its more tiered system, a sizable minority of products reach the 10% referral fee bracket, which was used to define the high-fee group in this binary variable. This right-skewed distribution implies heterogeneity in the sample.

Price levels differ considerably across markets, but some general trends are found. To facilitate interpretability, we present the raw Buy Box price in our descriptive statistics table and plot the log-transformed Buy Box price in Appendix Figure 3a. From the table, we observe the largest standard deviations compared to all the other indicators for all three countries, indicating substantial skewness in the Buy Box price distribution. From the graph, the log-transformed Buy Box Prices provide a more comprehensive picture: Japan's distribution is centered at a higher log price (mean ≈ 7.9), whereas the U.S. and France cluster around mean ≈ 3.0. This variation justifies the use of a centered log price term and separate country-specific specifications.

The 90-day average sales rank in millions is naturally right-skewed because the data collection targets the top-selling 5,000 products per subcategory. To facilitate linear interpretation, we take the log transformation. The resulting distributions (Appendix Figure 3b) reveal positively skewed distributions with long tails, indicating that most products have relatively lower sales ranks, with a few outliers occupying higher ranks. France exhibits the most compressed distribution with a tighter concentration near top sales ranks, while the U.S. and Japan show slightly more dispersion.

`Review_Count_Thousand` is highly right-skewed, with most products receiving very few reviews and only a small share accumulating large volumes. Using log-scaled bins, the line plot highlights clear cross-country differences: Japan has the lowest overall review activity, while products in the U.S. show higher average review counts and a longer right tail, suggesting a potentially longer product lifecycle or the culture of leaving reviews when online shopping.

In contrast, the `Avg_Review_Rating` shows a bell-shaped distribution across countries, centered around 4.2–4.5 out of 5. Japan's distribution skews slightly lower, suggesting more critical consumer feedback. The rating variable is useful for capturing perceived quality, but its limited variance (bounded between 1 and 5) may reduce predictive power.

For offer counts categories, monopoly listings dominate all three markets, especially in Japan where over 70% of products have only one seller. The U.S. and France exhibit more listings in the Low and Moderate categories. This distribution reflects broader differences in marketplace maturity and seller participation.

Out-of-stock frequency over the past 90 days is another key supply-side indicator. As Figure 5b illustrates, most products (>60%) are always in stock across three countries.

5 Methodology

In this thesis, we use a logit model to empirically investigate the probability that Amazon wins the Buy Box ($Y = 1$) versus a non-Amazon, third-party seller wins the Buy Box ($Y = 0$).

The logit model is appropriate for this setting for several reasons: Firstly, the outcome is binary, and we are interested in modeling the probability that Amazon wins the Buy Box, conditional on the explanatory variables we choose. Secondly, an alternative option, the probit model, relies on the assumption of normal distribution (Wooldridge, 2019). According to the descriptive statistics in Table 2, the distribution of the binary outcome is skewed toward third-party sellers, which supports the use of the logit model over a probit model, because the logit model assumes a logistic distribution (Wooldridge, 2019) and is more robust to skewed data.

Formally, the logit model estimates the following probability:

$$\mathrm{P}(Y = 1 \mid X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k)}} \tag{1}$$

This equation represents the probability that Amazon wins the Buy Box for a given product with characteristics X, where β_0 is the intercept and β are coefficients on the explanatory variables. The logit model can also be interpreted through a latent variable framework:

$$Y^* = \beta_0 + X\beta + \varepsilon$$

$$Y = \begin{cases} 1 & \text{if } Y^* > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{where } \varepsilon \sim \text{Logistic}(0,1)$$

Here, Y^* represents an unobserved utility or incentive that reflects Amazon's decision to allocate the Buy Box. The error term ε is assumed to follow a standard logistic distribution.

Because logit coefficients are expressed in terms of log-odds, they are not easily interpretable. Therefore, I compute Average Marginal Effects (AMEs), which represent the average change in the predicted probability that Amazon wins the Buy Box for a one-unit change in each explanatory variable, holding all others constant and averaging across all observations.

AMEs are especially useful in non-linear models like logit, where the marginal effect of a variable may vary depending on the values of other covariates. This approach also facilitates interpretation when interaction terms are included. For instance, in the interaction between price and referral fee, AMEs allow me to estimate the independent contribution of referral fees while still accounting for their interaction with price.

Because AMEs are non-linear transformations of estimated parameters, and several key variables are either categorical or involved in interactions, I compute robust standard errors using non-parametric bootstrapping. This method does not rely on asymptotic normality and provides reliable standard errors and confidence intervals. Specifically, I bootstrap the AMEs using 500 replications.

Finally, I interpret each model separately by country, rather than pooling all observations into a single regression. This allows me to identify country-specific patterns in Buy Box allocation and to avoid heterogeneity effectively.

6 Results

6.1 Regression Model Setup

We use a standard logistic regression model to estimate Amazon's probability of winning the Buy Box, as shown in Equation (1). The full list of explanatory variables used in our core regression model and their coding can be found in Appendix Table A.2.

To interpret the effects of each variable on the probability scale, we computed Average Marginal Effects (AME). This approach allows us to quantify the average change in the predicted probability of Amazon winning the Buy Box for a one-unit change in each explanatory variable, holding all other variables at their observed values. By leveraging AME, we address interpretability concerns associated with interaction terms in logit models and provide more intuitive, policy-relevant insights into Amazon's platform behavior.

6.2 Main Variable Interpretations

The AME results across the three countries, reported in Table 6 – 8, reveal a consistent relationship between referral fee levels and Amazon's likelihood of participating directly in the Buy Box. Using bootstrap-based marginal effect estimates, we find that in all three country-level logit models, the average marginal effect of referral fees is positive and statistically significant.

The result suggests that when the referral fee binary variable increases from 0 to 1, meaning that when a product changes from a low referral fee to a high referral fee category, the probability that Amazon wins the Buy Box increases by 17.94 percentage points in the U.S. (bootstrap SE = 0.023), 13.16 percentage points in France (bootstrap SE = 0.066), and 2.02 percentage points in Japan (bootstrap SE = 0.0089). These results are derived from marginal effects that account for interaction terms in the underlying logit models and are robust to resampling variation.

This finding contradicts our original hypothesis. Rather than maximizing commission income by encouraging third-party participation in high-fee categories, Amazon appears to prefer self-selling in these categories. One interpretation is that Amazon's behavior reflects a strategic decision to become a dominant player in high-frequency and essential categories, such as grocery. Even if this means forgoing higher commission revenue, Amazon may prioritize Buy Box control in these categories to strengthen customer acquisition and increase long-term market share. This behavior remains consistent with the broader concept of algorithmic self-preferencing: Amazon's incentives to self-preference may be driven not only by short-term margins, but by the strategic positioning of certain categories within its retail ecosystem. As a result, such self-preferencing in grocery may reduce the effectiveness of margin-based competition and deter third-party participation (Crémer et al., 2019).

6.3 Other Main Specifications

Besides referral fees, several other variables show consistent patterns across countries. In all three markets, Amazon is significantly less likely to win the Buy Box for higher-priced products. A one-unit increase in the centered log of Buy Box price, which is equivalent to approximately 2.72 times the average Buy Box price, reduces Amazon's probability of winning the Buy Box by 10.12pp in the U.S. (bootstrap SE = 0.0079), 8.41pp in France (bootstrap SE = 0.0194), and 5.22pp in Japan (bootstrap SE = 0.0057). This shows Amazon's emphasis on price competitiveness in grocery markets. Similarly, sales rank is strongly negatively correlated with Buy Box success. A one-unit increase in the log 90-day average sales rank, which is equivalent to the product's sales rank becoming approximately 2.7 times higher, reduces Amazon's probability of winning by 63.24pp in the U.S. (bootstrap SE = 0.051), 29.93pp in France (bootstrap SE = 0.229), and 27.94pp in Japan (bootstrap SE = 0.035), suggesting Amazon consistently favors grocery products with better sales performance.

Despite these consistencies, notable cross-country differences arise. For example, Amazon is more likely to win the Buy Box when the Buy Box is awarded to the lowest-price seller in the U.S., but the opposite direction of association is observed in France and Japan. When the Buy Box is awarded to the lowest-price seller, Amazon's chance of winning Buy Box increases by 2.46pp in the U.S. (bootstrap SE = 0.012), but decreases by 19.42pp in France (bootstrap SE = 0.229) and by 10.60pp in Japan (bootstrap SE = 0.0157). These results do not necessarily imply self-preferencing, but they reflect how Amazon's Buy Box outcomes differ when responding to long-term price competition in different geographic areas. In the U.S., Amazon appears to perform better when the market emphasis on long-term price competitiveness is stronger. However, in France and Japan, the platform appears less likely to secure the Buy Box under such conditions. This could suggest regional differences in Amazon's algorithmic strategy or the presence of stronger third-party competition in France and Japan.

Stock availability plays a stronger role outside the U.S. While the out-of-stock rate has no significant effect in the U.S., a one-unit increase in out-of-stock category (e.g., moving from always in stock category to low out-of-stock category), significantly reduces Amazon's likelihood of winning the Buy Box by 3.08pp in France (bootstrap SE = 0.0107) and 2.40pp in Japan (bootstrap SE = 0.0038). These results suggest that Japanese and European consumers may place greater weight on fulfillment consistency and reliability. Review quality also shows regional difference. In the U.S., both quality metrics are statistically significant, with one-unit increase in average review ratings increases Amazon's Buy Box probability by 7.62pp (bootstrap SE = 0.0112), and a one thousand increase in review count increases it by 1.07pp (bootstrap SE = 0.0017). In France, a one thousand increase in review count increases Amazon's probability by 1.61pp (bootstrap SE = 0.0042), while review rating is not significant. In Japan, neither review metrics significantly impacts Buy Box outcomes. These findings indicate that quality signals affect Buy Box algorithms differently across cultural contexts.

Competition intensity also shows cross-country variation in its association with Amazon's Buy Box outcomes. In the United States, a one-unit increase in the offer count category (e.g., moving from monopoly to low competition with 2-3 sellers), decreases Amazon's probability of winning the Buy Box by 0.74pp (bootstrap SE = -0.0077), but only being statistically significant at the 10% level. In Japan, the relationship is similarly negative, with a one-unit increase in the offer count category associated with a decreased probability of Amazon winning the Buy Box by 0.85pp but more statistically significant (p-val = 0.0036). However, in France, competition intensity is not statistically significant. The direction of association supports our hypothesis that higher competition reduces Amazon's probability of winning Buy Box in the U.S. and Japan, but the lack of a significant relationship in France could suggest different competitive dynamics.

6.4 Interaction Terms

Importantly, the interaction effects reveal how Amazon's Buy Box strategy changes under different economic conditions. We go back to the logit regression result table and calculated the coefficient for $\beta_2 + \beta_9$. In the U.S. market, the combined coefficient of price and its interaction with the referral fee is approximately zero ($-3.5381 + 3.5443 \approx 0.006$), suggesting that Amazon's likelihood of winning the Buy Box becomes insensitive to price in high-referral-fee categories. This means that even when Amazon's price deviates from the average, its probability of winning the Buy Box does not decrease, which is a sharp contrast to the strong negative price effect observed in the overall sample.

In France and Japan, the effect is similar but less pronounced. The combined coefficients are 0.048 and -0.492, respectively. This indicates the negative association between price and Amazon's probability of winning Buy Box in these two countries is partially mitigated in high-referral-fee categories. This result reflects differences in platform strategy, regulatory pressure, or marketplace structure. For example, France's 2022 EU regulatory commitments may have some effects on Amazon's pricing strategies; the Japanese online grocery market is more fragmented, as Amazon faces strong competition from platforms like Rakuten and AEON (TraceData Research, 2024). A more fragmented e-commerce landscape could create stronger external pressures for Amazon to remain price competitive across both high- and low-referral-fee categories in order to remain attractive as a platform.

The mitigated negative effect of price in high-referral-fee categories illustrates the Amazon's Buy Box allocation is shaped by strategic incentives rather than purely by price competitiveness. In high-referral-fee categories, Amazon earns revenue either by self-selling and capturing the full product margin or by earning higher commissions from third-party sellers. This dual revenue stream reduces Amazon's need to compete aggressively on price, leading to greater price flexibility when allocating the Buy Box. In other words, in high-referral-fee categories, changes in price have a significantly weaker effect on Amazon's probability of winning the Buy Box, reflecting an internal trade-off between commission income and direct sales profit. Importantly, this pricing flexibility is not available to third-party sellers, who must still pay referral fees and rely on competitive pricing to win Buy Box.

7 Robustness

7.1 Self-preferencing Indicator

One limitation of the current approach is the inability to isolate self-preferencing behavior using the `BB_Winner_is_LP_Seller` variable alone. While this variable captures whether the Buy Box winner was the lowest-price seller over the past 90 days, it does not specify the identity of the lowest-price seller. For example, Amazon may win the Buy Box while a third-party FBA seller offers a lower price, and that could be considered as a potential signal of self-preferencing. However, the same variable would also equal zero if Amazon was the lowest-price seller and a third-party seller won the Buy Box, which would suggest the opposite of self-preferencing. Without knowing the identity of the lowest-price seller and the Buy Box winner simultaneously, it is impossible to attribute observed patterns directly to self-preference.

To address this and boost our result's robustness, we construct a variable called `Self_Preference_Indicator`. It equals 1 when Amazon wins the Buy Box and is *not* the lowest price seller over the past 90 days, and 0 otherwise (i.e., `Amazon_Binary` = 1 and `BB_is_LP_Seller` = 0). We then calculated the mean frequency of self-preferencing by country and by referral fee categories. The results are in Table 9.

While the regression results show Amazon is more likely to win the Buy Box in high-referral-fee categories overall, our robustness check reveals that more questionable Buy Box wins are more frequent within low-referral-fee categories. Specifically, the rate at which Amazon wins the Buy Box despite not being the lowest-price seller over the past 90 days is higher within the low-referral-fee group compared to the high-referral-fee group. This suggests that, although Amazon generally prefers to sell in the high-referral fee category, it may engage in more aggressive algorithmic self-preferencing in the low-referral fee category.

These findings provide a more comprehensive understanding of Amazon's Buy Box allocation strategy across different margin environments. In high-referral-fee categories, Amazon tends to win more often, but those wins are more likely to be price-aligned, reflecting fairer outcomes under competitive logic. This behavior may reflect not only margin-based incentives but also strategic priorities such as customer acquisition, retention, and market share expansion. In contrast, in low-referral-fee categories, where third-party commission revenue is less valuable, Amazon appears more willing to favor itself even in the absence of a price advantage.

Together, these results reveal a dual mechanism: Amazon demonstrates broad but competitively justifiable dominance in high-margin categories, while implementing greater algorithmic favoritism in low-margin contexts. This pattern refines our understanding of platform strategy and illustrates how Amazon balances profit optimization with strategic category control.

One of the most surprising findings from this robustness check is that France exhibits the highest rate of potentially unfair Buy Box wins. This contradicts our initial hypothesis that self-preferencing would be weakest in France due to the European Commission's earlier digital platform regulations. The discrepancy suggests that formal regulation does not necessarily eliminate subtle algorithmic advantages. It is possible that EC enforcement has concentrated more on issues like data sharing and search ranking transparency, while pricing-based favoritism in Buy Box allocation remains underexamined. Additionally, the lag between regulation and algorithmic implementation may explain the persistence of such patterns. These findings underscore the importance of ongoing empirical monitoring, even in tightly regulated markets, and point to the ways algorithmic self-preferencing can persist through less visible ways, particularly in strategically important but less scrutinized categories like grocery.

7.2 LASSO and Different Specifications

To validate the robustness of our findings and address potential multicollinearity among predictors, we applied a Least Absolute Shrinkage and Selection Operator (LASSO) regression alongside the standard logit models. The LASSO model performs both variable selection and regularization by penalizing the absolute size of regression coefficients. Prior to model training, all variables were standardized using the StandardScaler function from the scikit-learn library in Python. This ensures that each variable contributes equally to the penalty term, which is crucial for fair coefficient comparison in LASSO regularization. The data was then split into 80% training and 20% testing subsets to support model generalizability.

LASSO was applied with an alpha value of 0.1, allowing for moderate regularization.

As a result, the key variables retained by LASSO largely align with those identified as significant in the logit models, reinforcing the credibility of our core specifications and interpretations.

`Referral_Fee_Binary` was selected with positive coefficients in all countries, which aligns with our main finding that higher referral fees are associated with a greater likelihood of Amazon winning the Buy Box. This consistency across both modeling approaches supports the argument that Amazon tends to self-sell in high-referral-fee categories, potentially to internalize margins rather than incentivize third-party sellers.

`Centered_Buy_Box_Price` was consistently selected with large negative coefficients, also confirming our logit finding that Amazon is less likely to win the Buy Box for higher-priced listings. LASSO also retained the interaction between centered price and referral fee in all three markets, with decreasing coefficients across countries (U.S. 1.63, France 1.43, Japan 0.93). This again, increases robustness for our logit result that Amazon becomes less responsive to price when allocating Buy Box in high-referral-fee categories. In particular, the near-zero combined effect in the U.S. model ($-3.54 + 3.54 \approx 0$) confirms that price no longer penalizes Amazon in high-referral-fee settings. This result is also supported by the LASSO results.

Other key variables selected by LASSO are consistent with the main regression insights. `Log_90_Sales_Rank` had negative coefficients in all countries (U.S. -0.91; France -0.35; Japan -0.86), similar to what we see in the main logit regression. `BB_Winner_is_LP_Seller` was retained across all markets, with a positive coefficient in the U.S. (0.09) and negative ones in France (-0.64) and Japan (-0.60), confirming the directional patterns found in our AME results. The role of review metrics also followed regional patterns. `Avg_Review_Rating` and its interaction with review count were retained only in the U.S. (0.44 and 0.61, respectively). In contrast, `Review_Count_Thousand` was selected in France (0.48) and Japan (0.09) but shrunk to zero in the U.S. `OOS_Categorized` had minimal effect in the U.S. (-0.0005), but was more negative in France (-0.36) and Japan (-0.60), consistent with our main results showing stock availability matters more in those markets.

Overall, the LASSO variable selection corroborates the main logit findings and further supports our interpretation of Amazon's strategic behavior in the Buy Box allocation process. The alignment across methods illustrates the robustness of our empirical approach.

To further validate these findings, we tested multiple alternative model specifications beyond the core model, including variations in the transformation of categorical variables (e.g., out-of-stock rate and offer count), exclusion of less predictive review metrics, and the use of centered log Buy Box price over the past 90 days. Across all specifications, the direction and statistical significance of key explanatory variables, such as referral fee binary, current Buy Box price, and their interaction, remained consistent. This robustness emphasizes the reliability of our conclusions.

8 Conclusion

8.1 Conclusion

This thesis provides empirical evidence on how Amazon's Buy Box seller allocation, shedding light on how platform economic incentives and different market contexts shape algorithmic outcomes. The key takeaways can be concluded as follows:

First, Amazon's probability of winning Buy Box is consistently higher in high-referral-fee categories across all three countries, contradicted with our initial hypothesis that Amazon would prefer to collect commission revenue from third-party sellers when referral fee is high. One of the possible explanations is that Amazon places strategic value on dominating grocery category, and the benefit to take long-term market control outweigh the short-term benefit of collecting referral fees from independent sellers.

Second, the effect of price on Amazon's probability of winning Buy Box is different between high and low referral fee categories. While a higher `Centered_Log_Buy_Box_Price` generally reduces Amazon's chances of winning the Buy Box across all three countries, this negative association is mitigated in high-referral-fee categories. Specifically, in the U.S., the interaction coefficient nearly cancels out the base price effect. This implies Amazon grants itself greater pricing flexibility in high-margin categories, a privilege third-party sellers do not enjoy since they have to pay higher referral fee by charging a higher price. Such strategic difference in Amazon's Buy Box algorithm design raises questions about platform neutrality and the extent to which Amazon adheres to fair competition principles as outlined in frameworks such as the EU antitrust rules (European Commission, n.d.) or proposed U.S. legislation(Hovenkamp, 2023).

Third, robustness checks using the `Self_Preference_Indicator` reveal that Amazon is more likely to win the Buy Box without being the lowest-priced seller in low-referral-fee categories. While Amazon's self-selling in high-fee categories may be motivated by strategic control, its Buy Box dominance in low-fee contexts appears more algorithmically driven and potentially harder to justify on competitive grounds. This insight motivates future work to include the `Self_Preference_Indicator` as a direct regressor in model estimation, providing a more formalized test of this effect.

Fourth, LASSO regression and different model specifications confirm the robustness of the thesis's core model choice. All key predictors show consistent directions of association and statistical significances as the core model across all different model specifications and all three countries.

Overall, the findings challenge the notion that Amazon's algorithmic decisions are either platform-neutral or driven solely by platform favoritism. Instead, they suggest that Amazon's Buy Box strategy is shaped not just by fair assessments of seller quality or direct self-preferencing, but also by deeper economic and strategic incentives. Therefore, it is important to interpret algorithmic outcomes within the broader context of platform objectives and profit-maximizing incentives.

8.2 Limitation

This thesis offers valuable insights into Amazon's Buy Box allocation. However, several limitations should be acknowledged.

First, the empirical strategy used by this thesis is observational, which limits our ability to make causal claims. Future research could adopt natural experiments, regulatory interventions, or algorithmic audits to identify causal relationships and better isolate the effects of platform incentives on Buy Box outcomes.

Second, this thesis conducts a cross-sectional analysis. This limits our ability to assess how Amazon's algorithmic behavior evolves in response to external pressures, such as regulatory enforcement or changes in market competition. A panel or time-series approach would enable the study of temporal dynamics, such as whether self-preferencing intensifies or diminishes following antitrust actions.

Third, this study focuses on the Grocery category, which is particularly relevant due to its referral fee structure and strategic importance to Amazon. However, generalizing these findings to other product categories should be done with caution, especially as many other categories (e.g., Books, Electronics) lack tiered referral fee structures. It is also important to acknowledge the mechanisms of self-preferencing may differ in different product categories where Amazon's commission structure is less variable.

Fourth, while the `Self_Preference_Indicator` provides a useful proxy for identifying potential algorithmic favoritism, it is derived directly from the dependent variable and was therefore used only in the robustness test for this thesis. Incorporating this indicator or a modified version into the core regression framework (e.g., as an outcome variable in a nested model) could enable more formal hypothesis testing around self-preferencing behavior and its economic incentives.

Finally, although referral fees play a significant role in revealing Amazon's profit-maximizing incentives beyond price margins, they do not fully capture the platform's commission revenues. Many additional forms of commission income, such as profits from Fulfilled by Amazon (FBA) services, and revenues from sponsored ads on the marketplace display pages, are not recorded in the product-level data available from Keepa. Future research could incorporate a more comprehensive set of commission revenue streams to deepen the understanding of Amazon's platform profits and the strategic incentives behind the Buy Box allocation.

8.3 Antitrust Implications

The findings of this thesis bring meaningful implications for antitrust enforcement in digital marketplaces. Existing regulatory frameworks, such as the European Commission's commitments with Amazon (European Commission, 2022), and Article 102 of the Treaty on the Functioning of the European Union(European Commission, n.d.), have largely adopted a one-size-fits-all approach to self-preferencing and market dominance. However, this thesis suggests that Amazon's Buy Box behavior is not uniform across different referral fee categories and appears to be shaped by underlying economic incentives. This highlights the need for a more context-specific regulatory response.

In low-referral-fee categories where algorithmic favoritism is more severe, regulatory attention should be directed toward enhancing transparency and ensuring that Buy Box allocation processes are not systematically disadvantaging third-party sellers without competitive justification. For example, a low tolerance threshold for Amazon winning the Buy Box despite not being the lowest-price seller may serve as a benchmark for identifying detrimental self-preferencing.

However, in high-referral-fee categories, Amazon's increased Buy Box success in these categories may reflect a strategic choice to prioritize self-selling and pursue long-term dominance, even at the expense of short-term commission revenue. In such contexts, traditional metrics like price competitiveness or isolated Buy Box outcomes may be insufficient to detect anticompetitive behavior. Instead, monitoring strategic self-preferencing requires a broader evaluative framework, such as tracking long-term evolution in Amazon's market share, or identifying persistent reductions in third-party visibility and diversity that might exist even when price appears neutral.

In summary, the thesis suggests that platform favoritism and exclusionary effects cannot be fully understood without reference to the incentive structures embedded in platform design. A more tailored and tier-sensitive policy is needed. This includes mandating disclosure of Buy Box win rates by referral fee tier, seller type, and relevant competitive metrics, accompanied by clear explanations of the algorithmic criteria used. Black-box audits could also be initiated when these disclosures reveal disproportionate outcomes.

References

Amazon. (n.d.). How much does it cost to sell with Amazon? https://sell.amazon.com/pricing#referral-fees

Amazon.co.jp. (n.d.). Selling Plans, Shipping Fees and Cost Calculator. https://sell.amazon.co.jp/en/pricing

Amazon.fr. (n.d.). Amazon Pricing and Amazon Seller Plans | Amazon FR. https://sell.amazon.fr/en/tarifs

Arbatskaya, M. (2007). Ordered Search. *The RAND Journal of Economics*, *38*(1), 119–126. http://www.jstor.org/stable/25046295

Armstrong, M., & Zhou, J. (2011). Paying for prominence. *The Economic Journal*, *121*(556), F368–F395. https://doi.org/10.1111/j.1468-0297.2011.02469.x

Buchi, S. (2025, March). Amazon Buy Box Guide. https://www.sarasanalytics.com/blog/amazon-buybox

Chen, L., Mislove, A., & Wilson, C. (2016, April). *An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace* (tech. rep.). https://doi.org/10.1145/2872427.2883089

Chen, N., & Tsai, H.-T. (2019). Steering via algorithmic recommendations. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3500407

Chen, Y., & He, C. (2011). Paid placement: advertising and search on the internet. *The Economic Journal*, *121*(556), F309–F328. https://doi.org/10.1111/j.1468-0297.2011.02466.x

Choi, H., & Mela, C. F. (2019). Monetizing online marketplaces. *Marketing Science*. https://doi.org/10.1287/mksc.2019.1197

Cialdini, R. B. (1993, January). *Influence: The Psychology of Persuasion*. http://cds.cern.ch/record/2010777

Crémer, J., de Montjoye, Y.-A., Schweitzer, H., & Commission, E. (2019). *Competition policy for the digital era* (tech. rep.). European Commission. https://doi.org/10.2763/407537

Duquesne, G., Bowman, T., de Bernard, T., Prasad, K., Armstrong, P., & Review, G. C. (2002, October). What constitutes self-preferencing and its proliferation in digital markets. https://globalcompetitionreview.com/guide/digital-markets-guide/fourth-edition/article/what-constitutes-self-preferencing-and-its-proliferation-in-digital-markets

Etumnu, C., & Noumir, A. (2023). The effect of Amazon Prime on sales ranks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4338961

Etumnu, C. E., & Malone, T. (2024). An empirical examination of BuyBox assignments in online grocery retail. *Agribusiness*. https://doi.org/10.1002/agr.21978

European Commission. (n.d.). Antitrust and Cartels overview. https://competition-policy.ec.europa.eu/antitrust-and-cartels/overview_en

European Commission. (2022, December). Antitrust: Commission accepts commitments by Amazon barring it from using marketplace seller data, and ensuring equal access to Buy Box and Prime. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7777

Farronato, C., Fradkin, A., & MacKay, A. (2023, January). *Self-Preferencing at Amazon: Evidence from Search Rankings* (tech. rep.). https://doi.org/10.3386/w30894

Federal Trade Commission. (2023, September). FTC Sues Amazon for Illegally Maintaining Monopoly Power. https://www.ftc.gov/news-events/news/press-releases/2023/09/ftc-sues-amazon-illegally-maintaining-monopoly-power

Hovenkamp, H. (2023). Antitrust and Self-Preferencing. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4526022

Hunold, M., Laitenberger, U., Thébaudin, G., of Siegen, U., i3, I. P. d. P., Telecom Paris, & CRED. (2022, June). *Bye-Box: An analysis of Non-Promotion on the Amazon Marketplace* (tech. rep.). https://cred.u-paris2.fr/sites/default/files/working_papers/CRED-WORKING-PAPER-2022-04.pdf

Jürgensmeier, L., Bischoff, J., & Skiera, B. (2024). Opportunities for self-preferencing in international online marketplaces. *International Marketing Review*. https://doi.org/10.1108/imr-03-2024-0094

Khan, L. (2017). Amazon's antitrust paradox. *SSRN Electronic Journal*. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2911742_code2291888.pdf?abstractid=2911742&mirid=1&type=2

Lill, M., Gallitz, N., Stich, L., & Spann, M. (2024, January). *Product Badges and Consumer Choice on Digital Platforms* (tech. rep.). https://doi.org/10.2139/ssrn.4935668

Long, F., & Amaldoss, W. (2024). Self-Preferencing in E-Commerce marketplaces: The role of sponsored advertising and private labels. *Marketing Science*, *43*(5), 925–952. https://doi.org/10.1287/mksc.2022.0262

Padilla, J., Perkins, J., & Piccolo, S. (2022). Self-Preferencing in Markets with Vertically Integrated Gatekeeper Platforms*. *Journal of Industrial Economics*, *70*(2), 371–395. https://doi.org/10.1111/joie.12287

Raval, D. (2023, December). *Steering in one click: Platform Self-Preferencing in the Amazon buy box* (tech. rep.). https://deveshraval.github.io/buyBox.pdf

Rory, V. L., & Aggarwal, N. (2023, May). Amazon's pricing paradox. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4436546

Rottembourg, J. M. B. (2024). Auditing the Ranking Strategy of a Marketplace 's Algorithm in the Frame of Competition Law Commitments with Surrogate Models: The Amazon 's Buy Box Case. *ideas.repec.org*. https://ideas.repec.org/p/gre/wpaper/2024-27.html

Shieh, G. (2011). Clarifying the role of mean centring in multicollinearity of interaction effects. *British Journal of Mathematical and Statistical Psychology*, *64*(3), 462–477. https://doi.org/10.1111/j.2044-8317.2010.02002.x

Statista. (2024). E-commerce in the United States | Statista. https://www.statista.com/study/28028/e-commerce-in-the-united-states-statista-dossier/

Statista. (2025, February). Top online stores worldwide 2024, by net sales. https://www.statista.com/forecasts/860716/top-online-stores-global-ecommercedb

The Japan Times. (2024). Antitrust watchdog raids Amazon Japan over alleged price-fixing. https://www.japantimes.co.jp/business/2024/11/26/companies/amazon-japan-raid/

TraceData Research. (2024). Japan Online Grocery Market Report, Market Size, Market sales, Market Share, Market Revenue, market Growth, Major players - TraceData Research. https://www.tracedataresearch.com/industry-report/japan-online-grocery-market

Tremblay, M. J., & of Business, F. S. (2021, August). *The limits of marketplace fee discrimination* (tech. rep. No. 20-10). https://www.tse-fr.eu/sites/default/files/TSE/documents/conf/2022/postal/tremblay.pdf

Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, *101*(3), 493. https://doi.org/10.2307/1885694

Wooldridge, J. M. (2019, January). *Introductory Econometrics: a Modern Approach*. Cengage Learning.

Appendix

A.1 Data Cleaning Process

A.1.1 Raw Data Processing

After collecting raw product-level data from Keepa across multiple subcategories, the first step was to combine all subcategories within each broader product category for each country. For example, all U.S. grocery subcategories were merged into a single dataset labeled `US_Grocery_Combined`, and the same approach was applied to France and Japan.

To prepare the data for analysis, we implemented a structured cleaning function to filter and standardize the key variables used in the regression. I began by selecting relevant columns, including `Buy Box Seller`, `Buy Box: Is FBA`, `Sales Rank: 90 days avg.`, `Buy Box : Current`, `Reviews: Rating`, `Reviews: Review Count`, `Buy Box: Unqualified`, `ASIN`, `Buy Box : 90 days OOS`, `New Offer Count: Current`, `Amazon: 90 days avg.`, `New, 3rd Party FBA: 90 days avg.`, `New, 3rd Party FBM : 90 days avg.`, and `Referral Fee based on current Buy Box price`.

Columns were renamed for clarity, and rows marked as `Buy Box Unqualified` were removed to ensure comparability across listings.

The variable `Avg_Review_Rating` was obtained directly from the raw Keepa dataset under the column titled Reviews: Rating, which reports the average customer star rating for each product at the time of data collection. Since this field was already pre-calculated by Keepa on a 1-to-5 scale, where 5 indicates the highest level of customer satisfaction, no transformation was required apart from standard renaming during the data cleaning process. However, in order to ensure data completeness, any observations with missing values for this column were excluded from the final regression dataset.

Since some price values used currency formatting (i.e., dollar symbols, and commas for decimal points), I implemented a custom conversion function to convert these text-based fields into numeric values. Rows with missing Buy Box prices were dropped, and the Buy Box seller was categorized into three fulfillment types: Amazon, FBA (Fulfilled by Amazon), and FBM (Fulfilled by Merchant). I also cleaned the 90-day OOS percentage column by removing the percentage symbol and converting values into numeric form for the next steps.

A.1.2 Binary Variable Cleaning

3 binary variables were constructed to support the logistic regression analysis. The primary dependent variable, `Amazon_Binary`, was coded as 1 if the Buy Box seller for a product was Amazon, and 0 otherwise (FBA or FBM). To evaluate long-run price competitiveness, I created the binary variable `BB_Winner_is_LP_Seller`, which equals 1 if the seller winning the Buy Box was also the lowest price seller over the past 90 days, and 0 otherwise. Lastly, a key explanatory variable for the Grocery category is `Referral_Fee_Binary`. Referral fees on Amazon are calculated as a percentage of the product's Buy Box price, but the exact fee structures differ across countries and product categories. For the United States, Amazon collects a referral fee of 8% of the Buy Box price for grocery products priced at \$15 or below, and 15% for those priced above \$15(Amazon, n.d.). Based on this policy, I classified products with a Buy Box price greater than \$15 USD as belonging to high-referral-fee categories (coded as 1), and those priced at or below \$15 as low-referral-fee categories (coded as 0). For France, the referral fee percentages follow the same pattern, but the price threshold is at €10(Amazon.fr, n.d.). Therefore, products with a Buy Box price above €10 were coded as high-fee (1), while those priced at or below €10 were categorized as low-fee (0). For Japan, the referral fee structure includes three tiers: 5% for grocery items priced at or below ¥750, 8% for those priced above ¥750 but at or below ¥1500, and 10% for items priced above ¥1500(Amazon.co.jp, n.d.). To ensure consistency with the binary classification used in the U.S. and France, I grouped the 5% and 8% tiers together as low-fee categories

(coded as 0) and assigned the 10% tier to the high-fee category (coded as 1).

A.1.3 Categorical Variable Cleaning

We created 2 categorical variables for regression. To preserve consistency in the modeling framework while adapting to local market characteristics, I ensured that the number of bins remained constant across countries and categories (e.g., 4 bins for `Offer_Count_Categorized`, 5 bins for `OOS_Categorized`, while allowing the bin thresholds to vary based on distributional differences.

For the number of offers, `Offer_Count_Categorized` is created. Most grocery products had between 1 and 5 sellers, with very few listings exceeding 10. Thus, I used a more compressed four-bin categorization for groceries: Monopoly (1 offer, coded as 1), Low Competition (2–3 offers, coded as 2), Moderate Competition (4–10 offers, coded as 3), and High Competition (11+ offers, coded as 4). This structure captured the competitive dynamics in a meaningful way for Grocery.

For the 90-day out-of-stock rate, `OOS_Categorized` is created. While many listings were in stock for the full 90 days, a substantial amount of Buy Box sellers have varied 90 days out-of-stock percentages. In the U.S., I observed a right-skewed but widely dispersed distribution. To reflect this range while maintaining interpretability, I created five OOS categories: 0% (always in stock, coded as 1), 1–7% (low OOS, coded as 2), 8–37% (moderate, coded as 3), 38–60% (high, coded as 4), and 61–100% (very high, coded as 5). I initially applied the same five-bin structure to JP and FR grocery as well, but visual inspection showed tighter distributions centered even more heavily around zero, especially in Japan. However, to preserve cross-country comparability and interpretive symmetry, I retained the five-level structure in both countries, using the same U.S.-based thresholds to allow direct comparison.

A.1.4 Scale Transformation

To improve interpretability and ensure consistency across product categories and countries, several continuous variables were scaled or transformed during the data preparation stage.

First, the 90-day average sales rank was log-transformed to variable `Log_90_Sales_Rank` to address right-skewed distribution and improve interpretability, as it captures the idea that the marginal effect of moving from rank 100 to 200 is different from moving from 10,000 to 10,100.

Second, to make the review count more interpretable and reduce the scale difference across variables, we divided the raw count by 1,000 and created the variable `Review_Count_Thousand`. A change in one unit of this variable now corresponds to an increase of 1,000 customer reviews, making the estimated marginal effects more practically meaningful when included in regression models.

Lastly, for the Buy Box price, we computed `Centered_Log_Buy_Box_Price`. We first apply a log transformation to Buy Box Price to address right-skewness and to allow for intuitive, percentage-based interpretations of pricing effects. Then, we centered this indicator by subtracting the sample mean of the log-transformed Buy Box price from each observation.

$$Centered\_Log\_Buy\_Box\_Price = \log(Buy\_Box\_Price) - \log(\overline{Buy\_Box\_Price}) \tag{2}$$

A one-unit increase in this variable implies that the Buy Box price is approximately 2.72 times higher than average, since $e^1 \approx 2.72$. This centering decision is motivated by both statistical and interpretive considerations.

First, centering is a well-established method for mitigating multicollinearity (Shieh, 2011). In this context, `Referral_Fee_Binary` is directly determined by thresholds in the Buy Box price. Therefore, including both variables in a regression model may introduce multicollinearity. Second, centering improves the interpretability of both main and interaction effects. After centering, the coefficient on the referral fee binary variable captures its average effect when the price is at the mean.

A.2 Table 1: Regression Variable Description

Table 1: Variable Descriptions and Coding

Variable	Description	Coding
Amazon Binary	Dependent variable for logit regression. Indicates whether the Buy Box winner is Amazon.	1 if the Buy Box Seller is Amazon 0 if the Buy Box Seller is a third-party seller (FBA or FBM)
Referral Fee Binary	Indicates whether a product falls into a high- or low-referral-fee category based on country-specific thresholds.	1 if the product is in a high-referral-fee category; 0 otherwise. Thresholds: • US: 15% if price > $15; 8% otherwise • FR: 15% if price > 10 euros; 8% otherwise • JP: 10% if price > 1500 yen; 8% if 751–1500 yen; 5% if ≤ 750 yen. Only 10% tier coded as 1
BB Winner is LP Seller	Indicates whether the Buy Box winner was the lowest price seller over the past 90 days.	1 if the seller type of the Buy Box winner matches the seller type of the lowest-price seller over the past 90 days; 0 if they differ
Centered Log Buy Box Price	Log-transformed Buy Box price centered by subtracting the sample mean.	Continuous
Log 90 Sales Rank	Log-transformed 90-day average sales rank.	Continuous
Offer Count Categorized	Number of competing sellers.	1: Monopoly (1 offer) 2: Low (2–3 offers) 3: Moderate (4–10 offers) 4: High (11+ offers)
OOS Categorized	90-day out-of-stock rate category (Grocery).	1: 0% (Always in stock) 2: 1–7% (Low) 3: 8–37% (Moderate) 4: 38–60% (High) 5: 61–100% (Very High)
Avg Review Rating	Average rating (1–5 scale).	Continuous
Review Count Thousand	Number of reviews (in thousands).	Continuous
Centered Log Buy Box Price × Referral Fee Binary	Interaction term for price and referral fees.	NA
Avg Review Rating × Review Count Thousand	Interaction term for average rating and review count.	NA

A.3 Descriptive Stats

A.3.1 Mean Values of Main Variables Across Countries

Table 2: Mean Values of Main Variables Across Countries

Variable	US	France	Japan
Amazon_Binary	0.15	0.17	0.06
Referral_Fee_Binary	0.73	0.84	0.77
BB_Winner_is_LP_Seller	0.88	0.89	0.94
Buy_Box_Price	32.68	29.04	4046.56
90_Sales_Rank_Million	0.19	0.08	0.22
Offer_Count_Categorized	1.74	1.60	1.43
OOS_Categorized	0.77	0.77	0.63
Avg_Review_Rating	4.26	4.22	4.07
Review_Count_Thousand	0.95	0.79	0.07

A.3.2 United States Descriptive Statistics

Table 3: United States Descriptive Statistics

	count	mean	std	min	max
Amazon_Binary	6040	0.15	0.36	0.00	1.00
Referral_Fee_Binary	6040	0.73	0.44	0.00	1.00
BB_Winner_is_LP_Seller	6040	0.88	0.33	0.00	1.00
Buy_Box_Price	6040	32.68	30.62	0.98	527.00
90_Sales_Rank_Million	6040	0.19	0.16	0.00	1.26
Offer_Count_Categorized	6040	1.74	0.90	1.00	4.00
OOS_Categorized	6040	0.77	1.18	0.00	4.00
Avg_Review_Rating	6040	4.26	0.64	1.00	5.00
Review_Count_Thousand	6040	0.95	4.54	0.00	95.16

A.3.3 France Descriptive Statistics

Table 4: France Descriptive Statistics

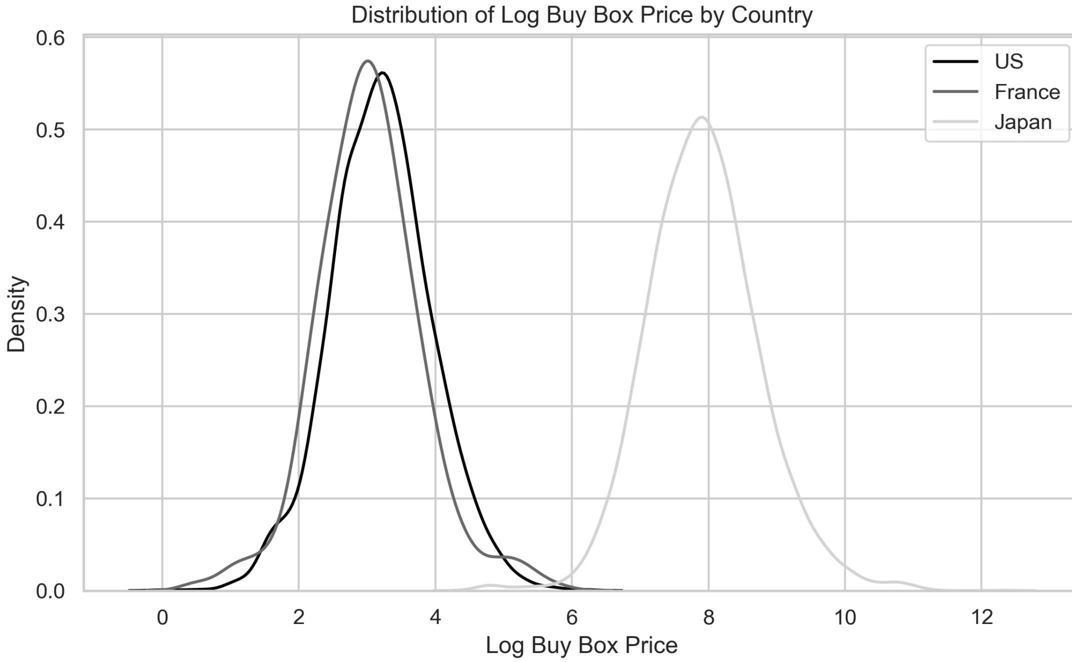
	count	mean	std	min	max
Amazon_Binary	1216	0.17	0.37	0.00	1.00
Referral_Fee_Binary	1216	0.84	0.37	0.00	1.00
BB_Winner_is_LP_Seller	1216	0.89	0.31	0.00	1.00
Buy_Box_Price	1216	29.04	34.10	1.55	368.80
90_Sales_Rank_Million	1216	0.08	0.18	0.00	2.73
Offer_Count_Categorized	1216	1.60	0.76	1.00	4.00
OOS_Categorized	1216	0.77	1.15	0.00	4.00
Avg_Review_Rating	1216	4.22	0.62	1.00	5.00
Review_Count_Thousand	1216	0.79	3.12	0.00	35.16

A.3.4 Japan Descriptive Statistics

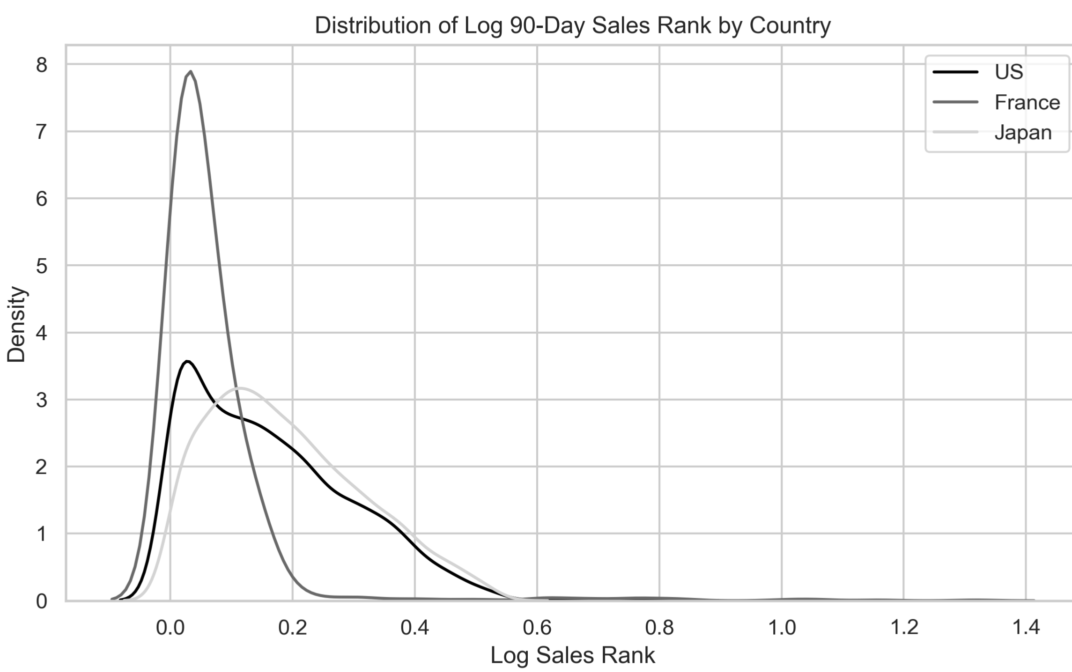
Table 5: Japan Descriptive Statistics

	count	mean	std	min	max
Amazon_Binary	7304	0.06	0.24	0.00	1.00
Referral_Fee_Binary	7304	0.77	0.42	0.00	1.00
BB_Winner_is_LP_Seller	7304	0.94	0.23	0.00	1.00
Buy_Box_Price	7304	4046.56	5595.02	88.00	215676.00
90_Sales_Rank_Million	7304	0.22	0.16	0.00	0.72
Offer_Count_Categorized	7304	1.43	0.74	1.00	4.00
OOS_Categorized	7304	0.63	1.06	0.00	4.00
Avg_Review_Rating	7304	4.07	0.85	1.00	5.00
Review_Count_Thousand	7304	0.07	0.70	0.00	36.71

A.3.5 Distributions of Key Variables by Country

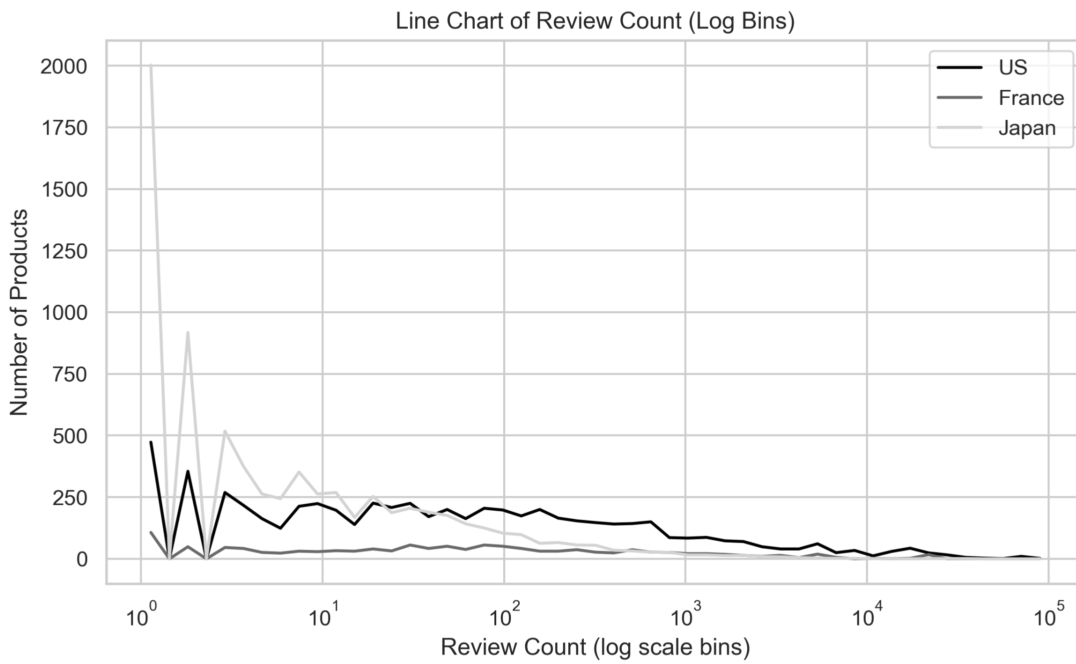


(a) Distribution of Log Buy Box Price

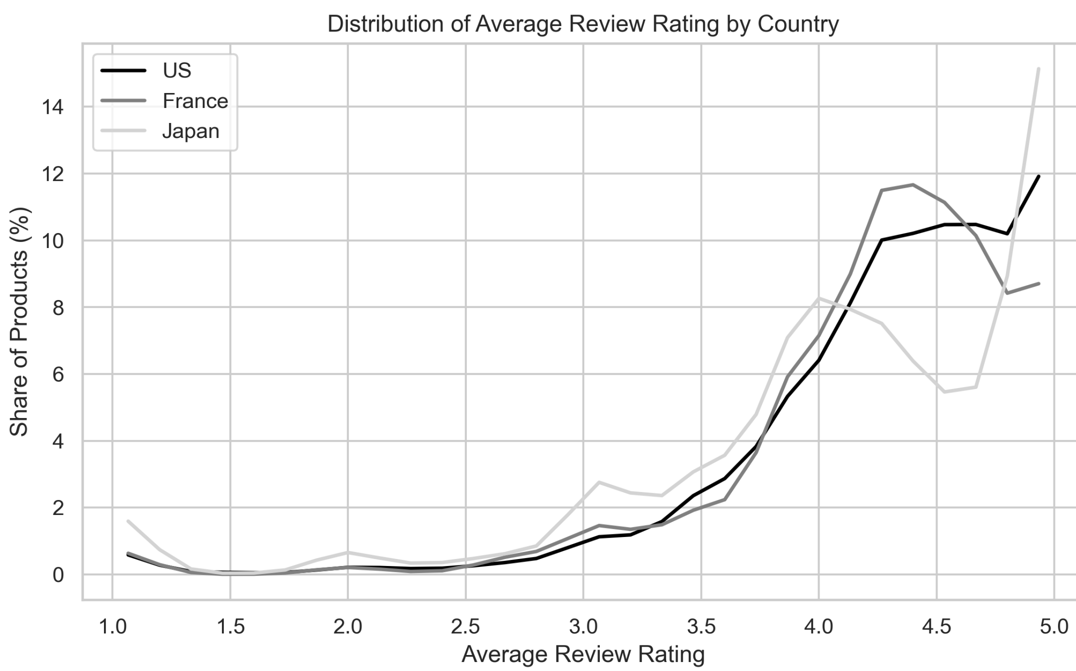


(b) Distribution of Log 90-Day Sales Rank

Figure 3: Price and Sales Rank Distributions by Country

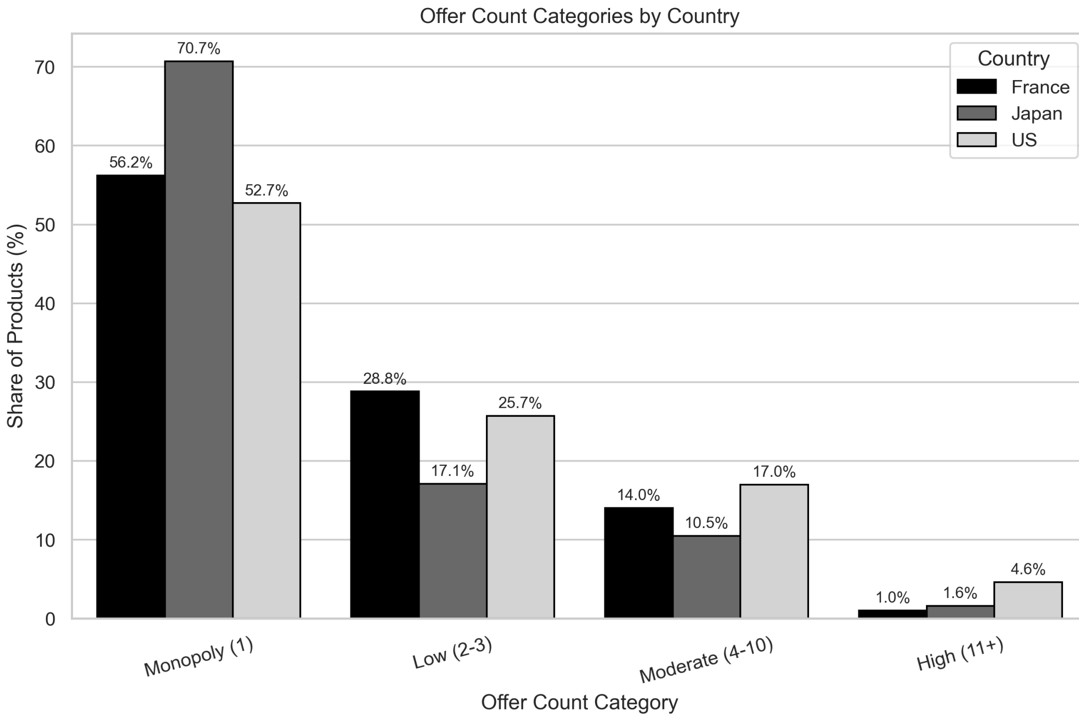


(a) Line Chart of Review Count)

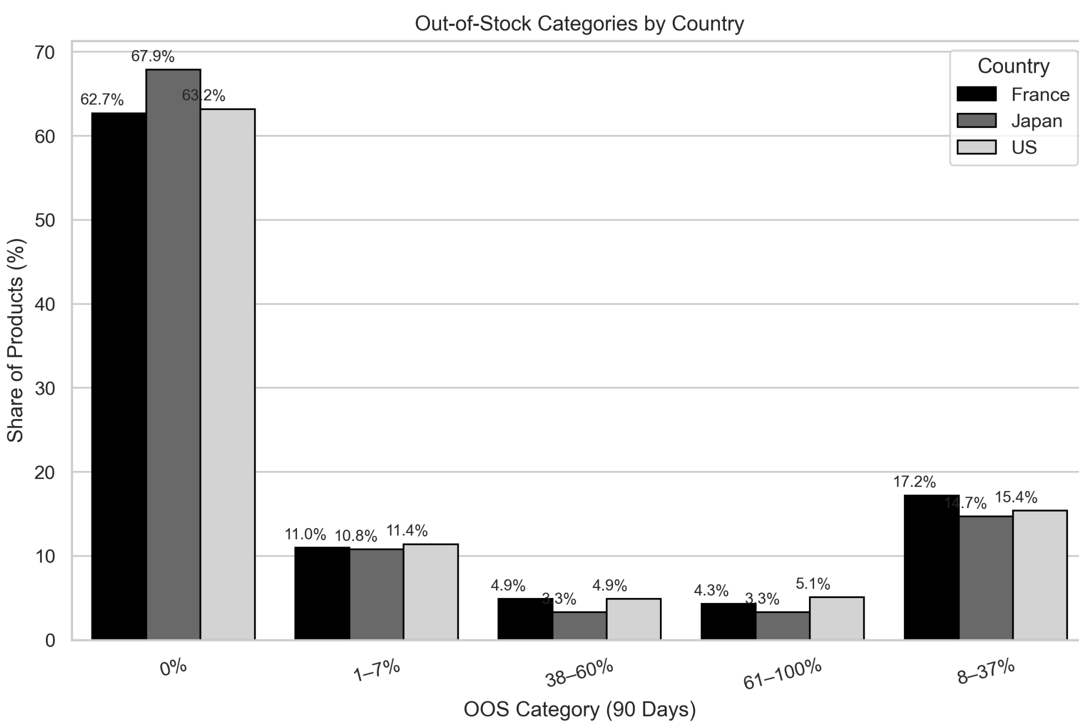


(b) Distribution of Average Review Rating

Figure 4: Review Distributions by Country



(a) Offer Count Categories by Country



(b) Out-of-Stock Categories by Country

Figure 5: Categorical Variable Distributions by Country

A.4 Regression Results

A.4.1 United States Regression

Table 6: United States Logistic Regression Results with AME

Variable	Coefficient	P-Value	AME (Point Estimate)	AME (Bootstrap Mean)	Bootstrap Std. Error
const	-6.6414	0.0000	N/A	N/A	N/A
Referral Fee Binary	2.4123	0.0000	0.1794	0.1822	0.0230
Centered Log Buy Box Price	-3.5381	0.0000	-0.1012	-0.1013	0.0079
BB Winner is LP Seller	0.2747	0.0402	0.0246	0.0255	0.0124
Log 90 Sales Rank	-7.0503	0.0000	-0.6324	-0.6367	0.0512
Offer Count Categorized	-0.0821	0.0867	-0.0074	-0.0077	0.0043
OOS Categorized	-0.0108	0.7835	-0.0010	-0.0011	0.0034
Avg Review Rating	0.6466	0.0000	0.0762	0.0768	0.0112
Review Count (k)	-0.6684	0.0121	0.0107	0.0107	0.0017
Price × Referral	3.5443	0.0000	N/A	N/A	N/A
Review Rating × Review Count	0.1795	0.0028	N/A	N/A	N/A

Note: Interaction term labels are abbreviated for clarity. *Price × Referral* refers to *Centered Log Buy Box Price × Referral Fee Binary*, *Review Rating × Review Count* refers to *Avg Review Rating × Review Count (Thousand). This holds for all the other countries' regression results.*

A.4.2 France Regression

Table 7: France Logistic Regression Results with AME

Variable	Coefficient	P-Value	AME (Point Estimate)	AME (Bootstrap Mean)	Bootstrap Std. Error
const	-2.1697	0.0340	N/A	N/A	N/A
Referral Fee Binary	1.8931	0.0060	0.1316	0.1404	0.0664
Centered Log Buy Box Price	-3.6147	0.0000	-0.0841	-0.0863	0.0194
BB Winner is LP Seller	-2.1390	0.0000	-0.1942	-0.1938	0.0189
Log 90 Sales Rank	-3.2964	0.0204	-0.2993	-0.3504	0.2289
Offer Count Categorized	0.0324	0.7946	0.0029	0.0028	0.0114
OOS Categorized	-0.3390	0.0010	-0.0308	-0.0319	0.0107
Avg Review Rating	0.0274	0.8700	-0.0185	-0.0210	0.0150
Review Count (k)	1.5011	0.0083	0.0161	0.0171	0.0042
Price $\times$ Referral	3.6634	0.0000	N/A	N/A	N/A
Review Rating $\times$ Review Count	-0.3133	0.0150	N/A	N/A	N/A

A.4.3 Japan Regression

Table 8: Japan Logistic Regression Results with AME

Variable	Coefficient	P-Value	AME (Point Estimate)	AME (Bootstrap Mean)	Bootstrap Std. Error
const	-0.7538	0.1120	N/A	N/A	N/A
Referral Fee Binary	1.3387	0.0000	0.0202	0.0207	0.0089
Centered Log Buy Box Price	-2.3407	0.0000	-0.0522	-0.0524	0.0057
BB Winner is LP Seller	-2.6043	0.0000	-0.1060	-0.1058	0.0062
Log 90 Sales Rank	-6.8647	0.0000	-0.2794	-0.2735	0.0351
Offer Count Categorized	-0.2090	0.0036	-0.0085	-0.0087	0.0031
OOS Categorized	-0.5891	0.0000	-0.0240	-0.0240	0.0038
Avg Review Rating	-0.0188	0.8305	-0.0031	-0.0040	0.0049
Review Count (k)	2.0058	0.1694	0.0133	0.0205	0.0157
Price × Referral	1.8495	0.0000	N/A	N/A	N/A
Review Rating × Review Count	-0.4120	0.1994	N/A	N/A	N/A

A.5 Robustness

A.5.1 Self-Preferencing Frequency

Table 9: Self-Preferencing Frequency by Country and Referral Fee Level

Country	**Overall**	**Low Referral**	**High Referral**
United States	1.7053%	2.4405%	1.4314%
France	5.4276%	13.6364%	3.8310%
Japan	2.1769%	3.6880%	1.7345%

Note. This table reports the percentage of Buy Box outcomes where Amazon wins the Buy Box despite not being the lowest price seller in the past 90 days. These cases are flagged by the `Self_Preference_Indicator` variable, which equals 1 when both `Amazon_Binary` = 1 and `BB_is_LP_Seller` = 0. Frequencies are shown overall and by referral fee tier.

A.5.2 LASSO Result

Table 10: LASSO Output Across Countries

Feature	**US**	**France**	**Japan**
Referral_ Fee_ Binary	0.9866	0.2655	0.4621
Centered_ Log_ Buy_ Box_ Price	-2.4844	-1.9180	-1.7661
BB_ Winner_ is_ LP_ Seller	0.0906	-0.6434	-0.6025
Log_ 90_ Sales_ Rank	-0.9061	-0.3548	-0.8626
Offer_ Count_ Categorized	-0.0576	0.0463	-0.1419
OOS_ Categorized	-0.0005	-0.3626	-0.6040
Avg_ Review_ Rating	0.4431	-0.0231	-0.0090
Review_ Count_ Thousand	0.0000	0.4823	0.0927
Centered_ Log_ Buy_ Box_ Price × *Referral_ Fee_ Binary*	1.6264	1.4320	0.9349
Avg_ Review_ Rating × *Review_ Count_ Thousand*	0.6110	0.0000	0.0000

Note. The table reports standardized LASSO coefficients from logit models estimated separately by country. Each coefficient reflects the relative importance of the corresponding feature in predicting Amazon's likelihood of winning the Buy Box, with higher absolute values indicating greater predictive power. A value of 0 indicates the variable was excluded by the LASSO penalty. All variables were standardized prior to estimation.