Fareed Khan April 10, 2024

Understanding Review Hijacking on Amazon Through Variation Listings: The ABCD's of The Bountiful Company

by

Fareed Khan

Maria Arbatskaya
Adviser

Economics

Maria Arbatskaya

Adviser

Samiran Banerjee

Committee Member

Marina Cooley

Committee Member

Ellen Gough

Committee Member

2024

Understanding Review Hijacking on Amazon Through Variation Listings: The ABCD's of The Bountiful Company

By

Fareed Khan

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 Honors

Economics

2024

Abstract

Understanding Review Hijacking on Amazon Through Variation Listings: The ABCD's of The Bountiful Company

By Fareed Khan

This thesis delves into the dynamics of product variants on Amazon's Best Seller Rank (BSR) within the Vitamins, Minerals, & Supplements category, leveraging a comprehensive dataset from Keepa.com spanning 6 months from November 2023 to February 2024. It specifically examines how factors such as the number of variants, average ratings, review volumes, pricing strategies, and participation in Amazon's Fulfillment by Amazon (FBA) and Subscribe & Save programs correlate with BSR improvements, employing statistical models like OLS, LASSO, and GLS for analysis. The findings highlight that a well-curated selection of product variants significantly bolsters BSR, while also pointing to the effectiveness of Subscribe & Save and FBA participation in strengthening market presence. However, the study identifies a threshold of 8 variants beyond which additional variants no longer contribute to BSR improvement, suggesting a balance is crucial to avoid consumer choice overload.

This thesis emphasizes the strategic manipulation of BSR through product variants by sellers, pointing to a gap in consumer awareness regarding the introduction of new variants to existing product listings. The thesis advocates for Amazon to implement clearer notifications for consumers when new variants are added, aiming to enhance transparency and support informed decision-making. This thesis not only sheds light on seller strategies in e-commerce but also brings to light the importance of safeguarding consumer interests through better marketplace regulations, contributing valuable insights into the interplay between seller tactics and consumer protection in the digital shopping realm.

Understanding Review Hijacking on Amazon Through Variation Listings: The ABCD's of The Bountiful Company

By

Fareed Khan

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 Honors

Economics

2024

Acknowledgements

I wish to start by extending my deepest gratitude to all the members of my committee for their invaluable support and dedication towards my academic journey. Dr. Gough, you spearheaded my passion for the liberal arts after my first year at Emory. Professor Cooley, you have been a highlight of my Goizueta experience, and your classes have engaged me intellectually more than any other class at GBS. I must express my profound appreciation to Dr. Arbatskaya, whose belief in my capabilities motivated me to elevate my thesis to its fullest potential.

Special thanks are due to Dr. Banerjee, whose guidance not only shaped the direction of this thesis but also played a pivotal role in inspiring me to pursue a major in Economics. Dr. Banerjee, your influence has been a guiding light, encouraging me to dedicate the last year to delve deeply into this topic.

I am immensely grateful to my parents and family, whose unwavering support over the last four years has been my foundation of strength. Their encouragement has been a constant presence, bolstering me through every challenge and achievement.

Finally, I cannot overlook the contribution of my best friend, Zayn Khan. Zayn, your expertise and advice on code has been instrumental in enhancing the impact of my thesis far beyond my initial expectations.

Contents

Contents

Appendix

Understanding Review Hijacking on Amazon Through Variation Listings: The ABCD's of The Bountiful Company

Fareed Khan

April 2024

1 Motivation

In this thesis, we investigate the phenomenon of review hijacking on major e-commerce platforms, an area with significant implications for global consumption yet lacking substantial research (Mayzlin 2014). The Federal Trade Commission's (FTC) actions, particularly under Chair Lina Khan, highlight the relevance of this topic. On May 19, 2022, the FTC proposed a new set of stringent guidelines aimed at combating the manipulation of online reviews—tactics that include the creation of artificial positive feedback and the deliberate suppression of negative reviews—indicating the escalating concern at the federal level (Federal Trade Commission, 2023).

Samuel Levine, Director of the FTC's Bureau of Consumer Protection, has raised alarms about the proliferation of deceptive practices, such as fake reviews (deliberately crafted false consumer feedback) and covert advertising (subtle product endorsements within content that appear to be unbiased) underscoring the gravity of these issues as they erode consumer trust and distort market competition (Federal Trade Commission, 2017). Per my personal conversation with the Chief of Staff for Advertising Practices on September 26, 2023, the

FTC has primarily caught cases of these through a combination of mediums including, whistle blowers, internal monitoring, and consumer tips.

The FTC's case against The Bountiful Company highlights these concerns. Founded in 1971, The Bountiful Company, a Nestlé-owned dietary supplements enterprise boasting 10 subsidiaries, generates over $6 billion in sales. One of the main subsidiary's health supplement products are widely sold across America under the label "Nature's Bounty." The case against The Bountiful revealed vulnerabilities in the e-commerce ecosystem, with them being charged for practices related to manipulating product reviews and endorsements. They were found to have misrepresented product ratings and endorsements, including claims about products receiving "Amazon's Choice" or "Best Seller" badges. The timeline of this case began in March 2020, when The Bountiful started engaging in review hijacking with their stress relief gummies. The e-commerce director at The Bountiful Company soon after in August 2020 disclosed a clever, albeit controversial, strategy: they aimed to propel new products into the limelight by creating variation relationships with established ones.

As months unfolded, The Bountiful Company put this strategy into action, introducing new products such as the Stress Comfort gummies and requesting they be varied with melatonin tablets. The impact was immediate and significant, with the director revealing that despite the Stress Comfort ratings being low initially, the variation strategy catapulted sales. By August 2020, they celebrated the strategy's success with the Zinc products, earning the coveted "Amazon's Choice" badge, a testament to their strategic ingenuity. They continued to engage in review hijacking actively throughout 2021 with a variety of products.

Figure 1: The Bountiful Company Review Hijacking Case Evidence

Note: This case of review hijacking is highlighted by the two products established as variants of each other. The Toy Story multivitamin was established as a variant for the

Vitamin C gummies. When products are listed as variants they share reviews with each other thereby aggregating the review count and the overall rating. However, these products are not the same as they are different vitamins.

Figure 1 showcases an example of the discussed review hijacking with the use of two dissimilar products. The products were listed as variants on the same product listing page, but although the description appears similar, they are different products. One of the products is a multivitamin while the other is a Vitamin C gummy. They were listed as size variants and were just one of the many cases of Bountiful being caught review hijacking. This was not merely an attempt to boost visibility; it was a calculated maneuver to amalgamate the ratings and reviews of new, less-known products with those of their well-received counterparts, thereby enhancing the overall appeal of the newer products.

However, what seemed like a masterstroke in digital marketing tactics soon came under scrutiny. By February 2023, the FTC took decisive action against The Bountiful Company, showcasing the manipulative side effects of their strategy. The climax of this tale was reached in March 2023 when the company agreed to pay $600,000 in redress, a move that not only marked the resolution of the case but also sent a clear message about the regulatory boundaries of online marketing practices (2023). Just recently, as of March 2024, the FTC announced that they would be giving out over $527,000 to consumers affected by The Bountiful Company Case if they submit a receipt showcasing that they were impacted by The Bountiful's use of review hijacking (2024).

The recent case involving The Bountiful Company has cast a spotlight on the pressing need for clear policy guidelines regarding the creation and presentation of product variants on e-commerce platforms like Amazon. This case exemplifies how the misuse of variant listings can lead to consumer deception and highlights the imperative for regulatory oversight in this domain. Abuses in variant creation not only distort consumer choices but also threaten the foundational trust upon which the digital economy is built.

To effectively regulate platform policies toward variant creation, it is essential for economists and policymakers to comprehensively understand the costs and benefits associated with variant creation from multiple perspectives. This understanding includes examining the incentives for sellers to create variants, assessing whether buyers genuinely benefit from an increased number of options, and evaluating how the interests of sellers, buyers, and the platform align or diverge.

This study takes a crucial step in this direction by empirically examining the impact of product variants on Amazon's Best Seller Rank (BSR). Our analysis reveals that Amazon's platform design currently incentivizes sellers to increase their number of variants, suggesting a strategic benefit for sellers in terms of visibility and sales potential. However, we identify a point of diminishing returns; the data suggests that beyond eight variants, the benefits may plateau or even decline. This phenomenon correlates with cognitive research suggesting that too many choices can overwhelm consumers, a concept formulated by Miller (1956) informally known as "the magic of 7", indicating a limit to the amount of information we can comfortably process.

So, what are the broader implications of these findings? This study not only points to the complex dynamics of variant creation on seller behavior and marketplace integrity but also serves as a foundation for policy recommendations. It points to the need for a balanced marketplace that optimizes the number of variants, catering to consumer convenience without overloading them, which in turn could inform Amazon's policy reforms according to Kinjo & Ebina (2014). By highlighting the nuanced impact of variant creation, this thesis lays the groundwork for future regulations that protect consumers from potential manipulations while still allowing sellers to innovate and differentiate their products effectively.

In the end, the goal is to craft a marketplace environment that strikes an optimal balance between variety and clarity, ensuring that the digital economy operates with both efficiency and integrity, and where consumer choice is empowered, not encumbered, by the abundance of options.

2 Literature Review

2.1 History of Review Hijacking

Review hijacking has evolved into a significant concern within the online retail space. Nicole Nguyen first brought the term to the public's attention in a 2018 BuzzFeed News article titled, "Here's A Guide To Spotting Fake Amazon Reviews." The term was coined to describe the manipulation of product reviews by third-party sellers on platforms like Amazon. This practice involves third-party sellers leveraging reviews from legitimate brand pages to sell counterfeit products. Prior to coining this term, Nguyen also discussed "review recycling," where sellers alter existing listings (title, description, and images) to capitalize on the accumulated reviews for entirely different products.

2.1.1 Defining Review Hijacking

Immediately following the BuzzFeed exposé, Review Meta (2018) officially defined "Review Hijacking." This definition included the practice of misappropriating reviews from one product to another, misleading consumers. This new definition focused on the appropriation of reviews rather than manipulation by outside parties. Review Meta is a metrics tool for evaluating the validity of Amazon reviews. They provide methods to detect such practices of review hijacking, citing examples like wheelchairs and Disney-themed paper towel, where reviews for these unrelated products were erroneously listed together. Review Meta accomplishes the regulation through their algorithm, which vets every review on a product's page and searches for descriptions unrelated to the product being sold. When errors are found, the reviews are flagged for the consumer so they're aware of possible hijacking.

2.1.2 Public Awareness and Consumer Reports Investigation

The discussion expanded when Consumer Reports published "Hijacked Reviews on Amazon Can Trick Shoppers" by Swearingen (2019). This piece highlighted how sellers manipulate

Amazon's review system, merging positive reviews from unrelated products to deceive consumers. Cases cited products like iPhone adapters and posture correction braces shared reviews for unrelated items on their product page, raising concerns about the reliability of review-based decision-making on e-commerce platforms.

2.1.3 Legal Actions and Continued Concerns

The issue of review hijacking gained further prominence with the legal case against The Bountiful Company, marking the U.S. Government's first legal action against such practices (2023). Despite efforts by Amazon and third-party analysts, review hijacking remains a pervasive problem, highlighting the need for more effective solutions to uphold digital marketplace integrity (Dayarni & Caverlee, 2021).

Review hijacking not only jeopardizes consumer trust but also challenges the integrity of online marketplaces. This necessitates ongoing vigilance and the development of robust mechanisms to maintain the reliability and usefulness of digital marketplaces for informed purchasing decisions.

2.2 Economic Implications of Review Manipulation

The intricate relationship between online reviews and consumer choices warrants a detailed exploration. A study by Mayzlin, Dover, and Chevalier (2012) is foundational in this regard, highlighting the strategic production of biased reviews by firms. Their thorough analysis using a difference-in-differences approach on data from TripAdvisor and Expedia provides critical insights. It reveals how factors like proximity to competitors and the physical size of a hotel influence the prevalence of fake reviews, thereby distorting consumer decision-making processes. They observed a 3.1 percentage point decrease in the gap between the proportion of 5-star reviews for chain hotels versus independent ones. This suggests a higher probability of smaller competitors resorting to posting positive fake reviews according to their study. Positive fake reviews refer to fabricated or falsely generated reviews that portray a product, service, or establishment in a favorable light, aiming to showcase a business or product in a favorable light. In the case of this study, the independent hotels utilized fake reviews in order to gain a competitive advantage over chain hotels.

Additionally, the work of Luca and Zervas (2015) offers a deeper dive into Yelp's ecosystem. Yelp is a social networking site and online directory that helps people find local businesses. Luca and Zervas examine the complex interplay between a business's reputation, competitive environment, and propensity to engage in review fraud. The authors conceptualize a model wherein a firm's product quality is categorized as high or low, and it makes strategic decisions regarding pricing and the level of bias in product reviews. This nuanced approach highlights the strategic interactions between pricing, review manipulation, and consumer perceptions. It emphasizes the significant impact review manipulation can have on consumer behavior in online marketplaces. This insight is crucial for understanding similar economic incentives and competitive pressures that might drive review hijacking on platforms like Amazon, influencing Best Seller Rank (BSR) and consumer behavior. Companies with products that have poor reviews and are not correctly priced could lead to a higher propensity to review hijack with their well-established products, mirroring what is shown on Yelp.

As we delve into the economic implications of review manipulation, it becomes crucial to consider the legal and regulatory framework that governs these practices.

2.3 Legal and Regulatory Context of Online Review Manipulation

The evolving legal and regulatory context, predominantly shaped by the actions and policies of the Federal Trade Commission (FTC), is imperative to this analysis. The FTC's enforcement strategies and guidelines provide a critical backdrop for understanding the complexities of online review manipulation. The Commission's role as a guardian of consumer protection and competition laws involves an administrative process governed by Section 5(b) of the FTC Act. This process, essential for addressing "unfair or deceptive acts or practices," involves issuing complaints, the potential for consent agreements, and public comment periods, illustrating the FTC's commitment to upholding market integrity and safeguarding consumer interests.

The FTC has made significant strides in strengthening advertising guidelines, particularly focusing on fake positive reviews and suppressing negative ones (2023). This regulatory tightening, especially regarding social media platforms, resonates deeply with this thesis core concerns. It highlights the FTC's commitment to maintaining the integrity of consumer feedback mechanisms, which significantly influence consumer purchasing behaviors and market dynamics.

Moreover, the FTC's proposed 2023 rule against illicit review and endorsement practices directly intersects with the central theme of this thesis. These practices distort consumer perceptions and undermine the trust essential to the functioning of online marketplaces. The Consumer Review Fairness Act (CRFA) (2017) and the FTC's revised Endorsement Guides (2023) further establish a legal framework that is highly pertinent to this research. The CRFA's protection of honest consumer opinions and the FTC's guidelines against manipulating consumer reviews offer a legal backdrop for analyzing the economic impacts of these deceptive practices. Additionally, the FTC's focus on the authenticity of social media influence, including measures against purchasing fake followers and subscribers, underlines the evolving challenges in digital markets. The FTC Endorsement Guidelines state, "Businesses would be prohibited from selling false indicators of social media influence, like fake followers or views. The proposed rule also would bar anyone from buying such indicators to misrepresent their importance for a commercial purpose (2023)." They make it clear that intent matters when purchasing followers and it's not necessarily banned; however, you cannot use them to exercise influence. This aspect is particularly relevant to this thesis, highlighting the need for robust detection and regulation systems in the digital marketing era. With this regulatory context established, we now turn our focus to how these concepts apply in the lens of marketing.

2.4 Marketing Impact of Deceptive Reviews

2.4.1 In-Depth Analysis of Deceptive Review Practices

Deceptive reviews significantly impact the world of online marketing. Pioneering studies by Mayzlin & Chevalier (2006) and Dellarocas (2006) provide a foundational understanding of the motivations and effects of fraudulent reviews on sales and product performance. Dellarocas' (2006) study highlights the role of online anonymity in posting positive fake praise of a company's products. The results of his study indicated that firms' manipulation of online forums can either increase or decrease value to consumers, depending on if manipulations align positively or negatively with product quality. Mayzlin & Chevalier (2006) use a differences-in-differences approach to analyze the impact of consumer reviews on book sales at Amazon.com and Barnesandnoble.com. They find that new favorable reviews at one site increase book sales relative to the other site, emphasizing the power of reviews in influencing consumer behavior.

2.4.2 Further Research Exploring Consumer Awareness

The extension of this discussion by Karabas et al. (2023) examines how review valence and consumer awareness of deceptive practices impact product perceptions and purchase intentions. Their findings indicate that when consumers perceive reviews to be manipulative, particularly in the case of overly positive reviews, their perceptions of the product and purchase intentions are notably affected. Karabas identifies the importance of authenticity in reviews for maintaining consumer trust. Consumers lose trust in reviews that overly lean positive but don't have the same views for excessively negative reviews, so it's essential for the company to maintain a reliable review page free of review hijacking.

2.4.3 Exploring Specific Cases

In a focused study on Amazon.com, He, Hollenbeck, and Proserpio (2021) investigate the market for fake product reviews. They find that a diverse array of products, including those with many reviews and high average ratings, engage in purchasing fake reviews. Fake review products lead to a temporary improvement in ratings and review counts but also result in a subsequent increase in one-star reviews, particularly for newer products, indicating a consumer backlash against perceived deception.

2.4.4 Utilizing Advanced Techniques of Detection

Moon et al. (2020) contribute significantly to identifying deceptive reviews by employing text pattern recognition and machine learning algorithms. Their approach to distinguishing genuine from fake reviews provides valuable insights into the characteristics of these reviews and the factors influencing individuals to engage in deceptive practices.

2.5 Brand Leverage In The World of Review Hijacking

2.5.1 Impact on Brand Communication

The study by Jeffrey K. Lee (2021) titled "Emotional Expression and Brand Status" investigates the dynamics of brand communication on social media. Lee's analysis of over 200,000 social media posts indicates a notable negative correlation between highly emotional communications and a brand's perceived status. This research is crucial in understanding how online reviews, a key component of digital brand communication, can influence consumer perceptions of brand status. In the context of review manipulation, reviews could be leveraged to reduce emotionality and improve the status of brands in the digital marketplace by companies deciding to delete reviews or create fake reviews that reduce overall emotionality.

2.5.2 Macroeconomic Implications

The macroeconomic study "The Condition of Economies" by Flisikowski and Kucharska (2018) explores the impact of global brands on national economies. They illustrate the significant correlation between brand value and economic development in countries where these brands are headquartered. The broader economic implications of deceptive online reviews are emphasized through their findings. The distortion of brand perception through these practices might not only affect individual companies but could also impact the economic stability of entire nations.

2.5.3 New Product Launch Strategies in the Era of Review Hijacking

In his comprehensive analysis, "New Product Launch Success: A Literature Review," Alexander Salmen (2021) discusses the factors contributing to the success of new product launches. Salmen emphasizes the importance of strategic communication and market alignment in successful product introductions, elucidating how these elements serve as facilitators and fundamental drivers of product acceptance and market penetration. The importance of strategic communication lies in its ability to effectively convey the unique value propositions and innovations of the new product to potential customers and stakeholders, crafting a compelling narrative that resonates with the market's evolving preferences and needs. Concurrently, market alignment involves meticulously calibrating the product's features, benefits, and positioning to match the specific demands and expectations of the target market segment. In the context of review hijacking, the distortion of consumer perceptions through manipulated reviews can significantly impact the success of new product launches. Product launches highlight the need for integrity and authenticity in online review practices, as they are integral to the effective launch and sustained success of new products in the digital marketplace.

3 Institutional Details for Amazon.com

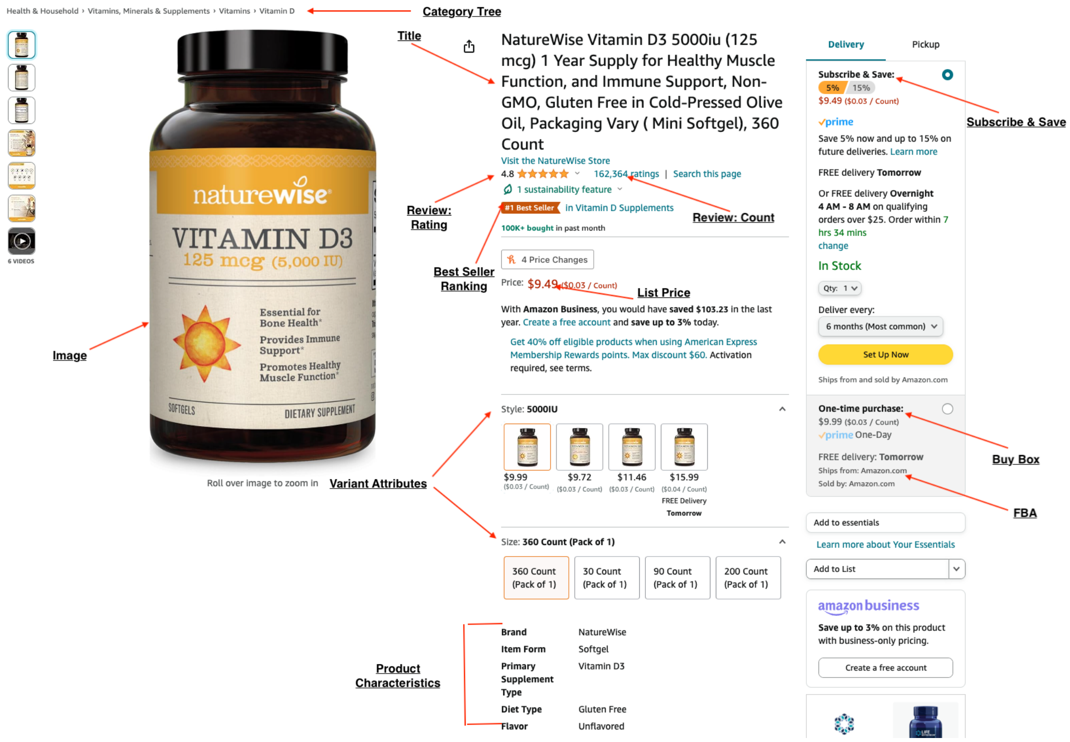
3.1 Connecting Variation Listings & BSR

3.1.1 Overview of Amazon's Variation Listing System

Amazon's variation listing system is a sophisticated mechanism that groups products into parent-child relationships. The parent listing, a non-buyable entity, serves as an umbrella for child products, which are the actual purchasable items. These child products vary in specific attributes like size, color, or flavor. This system plays a critical role in Amazon's marketplace, as it significantly influences how products are displayed and aids customers in comparing different options within a single listing. Understanding the dynamics of this system is crucial for analyzing Amazon's product performance. It affects the customer's browsing experience and has implications for sales and Best Seller Rank (BSR) dynamics. How products are grouped and displayed can impact consumer choice and, consequently, the sales performance of individual products.

To assist in understanding how these variables appear for a consumer when they are using Amazon on their day-to-day, we have included a product listing page along with annotations to showcase the different features in Figure 2. As the product page displays, the Best Seller Ranking, Review Ratings, # of Reviews, Price, along with other attributes are centralized in the middle of the page for the ease of the consumer's view. In addition, the purchasing area on the right side, showing prime eligibility, includes a Subscribe & Save feature in which one can purchase a product on a recurring basis for a discount. There also is the Buy Box feature which lies below this and you can find who fulfills the product (FBA or FBM). Lower on the page, we can find the various attributes showing the differing sizes along with styles which can vary greatly from product to product but this is just an example of a popular product page with multiple variant styles. These differing attributes are just a few of the many important ones you can find in the Amazon ecosystem.

Figure 2: Example of A Product Listing Page



3.1.2 Understanding BSR

The Best Seller Rank (BSR) is a significant metric on Amazon, indicating a product's sales performance relative to others in its category. Updated hourly, the BSR reflects a combination of recent and historical sales data along with proprietary Amazon data. Recent data refers to sales data from the past year while historical refers to sales over 1 year. Amazon's BSR system does not limit new products from receiving a high BSR as large volumes of short-term sales can push a product to the top of a category's sales rank. It is a critical indicator for consumers and sellers, as it signifies product popularity and success.

3.1.3 BSR's Influence on Seller Strategies

For sellers, achieving a high BSR is often a primary goal. This status can lead to increased visibility and sales. This further aligns with Skrovan's (2018) notions regarding the benefits a high BSR has on products on Amazon, explaining why many sellers are incentivized to improve their BSR. However, the pursuit of a high BSR can also lead to unethical practices such as review manipulation, emphasizing the need for vigilance and integrity in the digital marketplace. To fully grasp the economic implications of Amazon's systems, it is essential to combine an understanding of the variation listing system, search order ranking algorithms, and BSR dynamics. This comprehensive analysis reveals how these elements influence consumer behavior, product visibility, and, ultimately, sales performance on Amazon. For researchers and practitioners, this in-depth exploration provides critical insights into the mechanisms driving success on one of the world's largest e-commerce platforms. It underscores the importance of ethical practices, the impact of algorithmic biases, and the strategies sellers must navigate to succeed in this highly competitive digital marketplace.

3.1.4 Interplay Between BSR and Review Manipulation

The relationship between BSR and review manipulation is particularly noteworthy. Reviews can directly influence a product's sales, affecting its BSR. Practices like review hijacking, where sellers manipulate reviews to boost their products' perceived quality, can lead to artificial improvements in BSR. This manipulation distorts the metric's reliability as a measure of product quality and consumer preference.

3.1.5 Price Fluctuations Within BSR

Utilizing the Keepa API, our research entailed the creation of trend graphs for an in-depth analysis of numerical data variations across selected parent products and their variants, paralleling the analytical framework of He, Hollenbeck, and Proserpio (2021). The observation period spanned from the product's inception on Amazon to February 29, 2024. This comprehensive approach involved collating price data for variants and parent products within Vitamins A, B, C, and D, assembling four illustrative graphs in Figure 3. We pulled pricing data from the #1 Best Sellers in each respective vitamin category. These graphs showcase the absence of a unified pricing strategy across the variants in these categories as of

February 2024, highlighting the strategic use of pricing variations—either as loss leaders to boost immediate sales or through premium pricing to elevate profit margins. These findings reflect significant dips and spikes in the pricing graphs for specific variants where products are sold at a premium or at a discount. Notably, the price dynamics of the best sellers in Vitamins A and B demonstrate marked price fluctuations, pointing to a complex landscape of pricing strategies. We found it particularly interesting on the Vitamin D3 graph to see such extreme peaks in the blue line (B00GB85JR4) and the large troughs in the pink line (B093N7JJYL), further revealing a lack of a uniform pricing strategy. This inconsistency in pricing trends provokes a reassessment of pricing's impact on sales rank, interrogating its uniformity, context-dependence, or visibility primarily during price adjustments. Such findings prompt further inquiry into the strategic modeling of pricing as a pivotal element in securing a leading position on BSR.

3.2 Impact of Search Order Ranking on Product Visibility

3.2.1 Influence of Amazon's Search Algorithm

Research into Amazon's search ranking algorithms, such as studies by Sorokina & Cantú-Perez (2016), highlights the complex factors influencing product visibility on the platform. The study sheds light on the complex linkage of machine learning models, prioritizing consumer interaction factors, including intrinsic product properties, contextual features, and customer behaviors like purchases, clicks, and reviews. Amazon's transition to machine learning algorithms has highlighted their priority for behavioral data when curating search outcomes, showcasing the interaction between user engagement and algorithmic optimization. Although this approach to the ranking shows a commitment to search relevancy, there is an opportunity for many companies to leverage these tools, i.e., purchasing fake reviews, sponsoring products, etc., to improve search ranking.

3.2.2 Potential Biases and Implications

Recent studies from Etumnu (2022) suggest potential biases in Amazon's ranking system. Products enrolled in Amazon's Fulfillment by Amazon (FBA) program might receive preferential treatment in search rankings as the FBA program have a fixed fee for every product sold causing an increase in revenue for Amazon. His findings raise important questions about fair competition and the potential distortion of consumer choice. Further analysis reveals that FBA products have disproportionately ranked higher than those filled by merchants (FBM) with Etumnu's models indicating FBA having anywhere from a .3051% worsening to a .8307% improvement in BSR. His findings indicate a possible implicit preference that could mislead consumers and disadvantage independent sellers. Understanding these biases is critical for sellers who aim to optimize their product visibility and sales on the platform.

3.3 Comparative Analysis of The Amazon Badge System

3.3.1 Detailed Examination of Each Badge

The "Best Seller" badge, awarded based on sales volume within specific categories, reflects a combination of recent and historical sales. Though not fully disclosed, the criteria for this badge are understood to include sales velocity both in the short-term and long-term along with proprietary Amazon data. Products with this badge typically experience a substantial increase in visibility and traffic, as Skrovan (2018) reported. However, achieving this status can be challenging, and some companies engage in unethical practices like review hijacking to manipulate their way to the top.

In contrast, the "Amazon's Choice" badge, introduced in 2015, has evolved into a product quality and reliability symbol. This badge is contingent upon keyword relevance, customer ratings, product availability, and Prime eligibility. Products with high conversion rates for specific keywords and low return rates are typically favored. The badge significantly enhances product conversion rates, demonstrating its impact on consumer purchasing behavior.

3.3.2 Analyzing the Comparative Impact

While both badges substantially influence market visibility and conversion rates, their underlying criteria and resultant market impacts differ, as “Best Seller” is determined by long and short-term sales and has over 25,000 products with this badge out of over 4 Billion. Keyword relevance, customer ratings, and product and Prime eligibility determine “Amazon’s Choice.” Understanding these distinctions is crucial for sellers navigating the Amazon marketplace and strategizing to achieve these badges. This analysis of Amazon’s badges sets the stage for a deeper exploration of the marketing impact of deceptive reviews, mainly how they influence consumer perceptions and purchasing decisions.

4 Data

4.1 Data Overview

This study utilizes a comprehensive set of data from Keepa.com, focusing on the top 10,300 Best Sellers in the Vitamins, Minerals, & Supplements product category. Keepa.com is a tool for navigating the Amazon marketplace, offering downloadable datasets that track over 4 billion products across 12 countries. This vast repository provides information on price fluctuations, sales rankings, and market trends, enabling users to make informed decisions regarding their purchases. Keepa’s extensive coverage offers a global perspective on e-commerce dynamics in which you can find the top selling products for any category or subcategory. The data set focuses on a data set of Best Sellers from the past 6 months (September 2023, October 2023, November 2023, December 2023, January 2024, and February 2024). It provides a granular view of Amazon’s market dynamics, helping understand how products achieve the Best Seller badge. In addition, we pull yearly data for February 2024, 2023, 2022, and 2021. This was to analyze differences in summary statistics between each year and follow trends among the product category.

4.1.1 Data Collection Methodology

Through an analysis of Best Seller CSVs, we tailored the Python script to facilitate the extraction of data in the Vitamins, Minerals, & Supplements product category. to This allowed us to engage in time-series analysis for various variables, including variant counts across years.

4.1.2 Variables

The Keepa dataset includes several variables that help understand how products perform on Amazon. These variables include “Sales Rank: Current”, “Avg. Rating”, “Number of Reviews”, “Variation ASINs”, “Last Price Change”, “Age”, “Price”, “FBA”, and “Subscribe and Save”. The Sales Rank: Current variable indicates a product’s present position in Amazon’s sales ranking system, showing how well it’s selling compared to other products in the same category at the moment the data is pulled. The Avg. Rating shows what customers think of a product based on a scale, from zero to five stars, where a higher rating means better customer satisfaction. The number of reviews tells us how many customers have left feedback for a product, which can boost consumer confidence in reviews and ratings. Variation ASINs are unique codes assigned to each version of a product, helping to track and compare how different versions sell. The last price change shows when the price of a product was last updated, which can affect how often it sells if the price goes up or down. Age gives us an idea of how long a product has been available on Amazon.

Sales data, like how many were bought in the past month, gives a clear picture of how well a product is doing recently. The New: Price is the lowest cash value the seller asks for the consumer to pay for a product in a new condition. The FBA variable tells us if a product is being sold through Amazon’s Fulfillment by Amazon service, which can make products more attractive to buyers since they often come with faster shipping and Amazon’s customer service. Lastly, Subscribe and Save is a feature that lets customers sign up for regular deliveries of a product, often at a discount. This can make customers more likely to stick with a product since they get it automatically and might save money. Together, these variables give a detailed picture of a product's performance on Amazon, from how much people like it to how it is sold and delivered.

Table 1 provides a further description of the unique variables in the data set. The Keepa dataset provides a comprehensive analysis of the factors influencing BSR and further insights into how these factors can be manipulated to achieve BSR. Through this study, we will place significant importance on variants to assess their effect on a product's achievement of BSR. From this, we seek to discover the interplay of product performance, market dynamics, and consumer preferences on Amazon. Table 4 showcases the correlations between the variables via a correlation matrix.

4.2 Data Cleaning

Each data set includes 103 unique columns, including product identifiers, various sales ranks, pricing data, and categorical data related to the product classification. To begin the data cleaning, we remove columns that are not necessary to engage in further analysis, such as "Locale" (country of purchase) and "Title" (product name as listed on Amazon). These are both removed as they do not add to the regression analysis and provide a difficult means for cleaning the data to a sufficient degree. After this, we compiled all the data sets into one data frame to run a regression on the complete data set. To compile our data, we kept the ASIN (similar to a product identified) fixed starting in the month of September and added each subsequent data set for each month, ensuring the products stay fixed across time.

This data set provides a wide range of variables for analysis that describe various aspects of the products within the Vitamin, Minerals, & Supplements Category. To accelerate the process of data cleaning, we split the variables into categorical, numeric, binary, time, and percentage data as listed in Table 2. The categorization provided appropriate preprocessing techniques for each data type, enhancing the ability of the model to learn from the data effectively.

Categorical data, represented by features such as “Buy Box Seller”, “Category”, and “Brand”, undergo a one-hot encoding, transforming these categorical variables into a format that our model can interpret as dummies. One-hot encoding significantly increases the dimensionality of the dataset by creating a new binary column for each category level, which is particularly useful for models that struggle to interpret raw categorical data. Numeric data which includes a wide variety of product information, such as sales rank and pricing information, are converted from float to numeric types, with any errors or NaN values being dropped from the dataset so as to ensure the model runs without error. This step provided a consistent format for further statistical analysis. Furthermore, percentage data are converted into a numeric format by removing the percentage signs and dividing by 100 to standardize their values across variables and provide similar numbers for analysis. From there, all dollar signs were dropped as they didn’t provide relevant data for analysis and reduced the overall noise in an already dimensional data set. All time variables, except “Days Since Price Change” (which was converted to days), were converted into new variables for years since February 29, 2024 using the datetime package. Binary data such as “Subscribe & Save” and “Buy Box: Is FBA” are converted to 1s and 0s with 1 signifying a “yes” and 0 signifying a “no”. The new binary code provides a simplified means of testing against other variables and assessing their impact on the dependent variable.

To further the analysis of the effects of a variant relationship, a step involving the counting of the “Variation ASINs” column created a new variable, “Variation ASIN Count”. This variable represents the number of different variations (in size, color, or package quantity) that a single product ASIN (Amazon Standard Identification Number) has. To compute the “Variation ASIN Count” the dataset’s “Variation ASINs” column, initially filled with strings listing different ASINs separated by commas, was targeted. Due to the presence of products without any variations, missing values in this column were first filled with an empty string to ensure consistency. Subsequently, a split function was applied to this column to divide each string by commas and count the number of elements in the resulting list, thus providing the count of variation ASINs for each product. We also created a column in our dataset showcasing the various variant attributes for all products in our data set.

4.3 Descriptive Statistics & Preliminary Analysis

Table 3 includes the summary statistics of the variables we moved forward with testing on. From the earliest data we could obtain from Keepa, we tracked variation counts starting in February 2021 and continued to do so through February 2024. We completed this in the Vitamin subcategory and on the broader Health & Household category. After cleaning the data, we created line graphs of the average variation ASIN count with trend lines, as seen in Figure 4. In the Health & Household category, we found a trend in the increase in variants from February 2021 to February 2024. The trends continued in the broader Vitamin subcategory, where we saw a small but upward trajectory in the number of variants. The changes in the number of variants shed light on the continuous motivation for companies to add variants to their product listings. These stark changes are a clear trend in Amazon's broader Vitamins, Supplements, & Minerals market, showcasing possible continuities in The Bountiful's strategy across its competitors.

In a detailed examination of variant counts within our dataset over the recent six-month period, we introduced a novel variable titled "Has Variants." This variable is a binary indicator, distinguishing products based on the presence or absence of a parent-child relationship. After creating this variable, we found the percentage of products that do or don't have variants. The analysis, visualized in Figure 5, sheds light on the variant/multi-variant relationship within the Vitamin market, particularly revealing a relatively low percentage of products offering variants in the 6 months we analyze. Consequently, this aspect of product strategy warrants further investigation, particularly in understanding how variant offerings correlate with sales rank performance in the Vitamin market.

In our analysis of FBA, we explored Etumnu's (2022) hypothesis by constructing a box plot to compare BSRs between products fulfilled by Amazon and those that are not. The

"Sales Rank: Current" was log-transformed to standardize the y-axis, enhancing the clarity of our dataset visualization in Figure 6. The comparative analysis indicates that the median and mean sales rank of products utilizing Amazon's FBA service is significantly lower than that of non-FBA products, highlighting the trend of Amazon's fulfillment services on sales rank. These conclusions are corroborated by our Mann-Whitney U Test and t-test which both show significance at the 99% confidence level. The data implies that leveraging FBA could have a strategic advantage for achieving higher BSRs, encouraging sellers to consider paying to obtain the benefit of FBA.

5 Empirical Model

5.1 Specifications

In our empirical analysis, the Best Seller Ranking (BSR) is subjected to a natural logarithmic transformation to standardize the values within our data set. This transformation facilitates the reduction of our data's range and more accurately elucidates the subtle impacts of predictor variables on the sales dynamics within Amazon's marketplace. Also, this allows us to interpret the effects of RHS variables as the percentage change in BSR. Our empirical research proceeds with six methodologically distinct OLS specifications, each designed to systematically peel back the layers of the individual effects of each of the variables within our data set. OLS Regression serves as our foundational analytical tool. Its primary advantage lies in its simplicity and interpretability, providing a clear starting point for understanding relationships between our variables. OLS is particularly effective in cases where predictor variables are relatively independent of each other and the relationship with the outcome variable is linear. Given the direct impact we hypothesize that variant counts has on Best Seller Rank (BSR), OLS allows us to establish baseline effects and identify significant predictors within our dataset.

The initial specification establishes our preliminary analysis. It is an integrative model that encompasses a breadth of variables, including Subscribe & Save, FBA, Variation Count, Variation Count Sqrd (to capture potentially non-linear effects), Avg Rating, Number of Reviews, Age, Price, # Bought In Past Month, and Days Since Price Change. We control for month and types of Vitamin A, B, C, & D variables. Progressing to the second specification, we purposefully eliminate category controls. This refinement is crucial for isolating and understanding the month-specific fluctuations that inform purchasing behavior.

The third specification functions as a precision instrument, methodically removing the price variable to assess its underlying impact, yet preserving the extensive controls for both time and product category. This analytical process is replicated and intensified in the fourth specification, which isolates the temporal dimension to scrutinize the impact of price omission, thereby separating the influence of seasonality from that of category-specific factors.

The penultimate specification, the fifth, removes the variable indicating the # Bought In The Past Month, thus enabling an analysis of the residual variables within the enduring framework of month and category controls. Finally, the sixth specification refines the analysis to its essential elements, maintaining solely temporal controls to offer an unadulterated perspective on the seasonal patterns that choreograph the BSR.

Going deeper into what influenced our decision to include each variable, the Avg. Rating variable indicates the overall satisfaction of customers with a product based on the ratings they provide. A higher average rating typically suggests better customer satisfaction. We hypothesize that products with higher average ratings will generally have better BSR due to perceived quality and customer satisfaction. The # of Reviews variable reflects how engaged customers are with a product, as it measures the total count of reviews a product receives. More reviews can imply more trust. From preliminary research into fake reviews by Mayzlin & Chevalier (2006), we believe that products with a higher number of reviews tend to have a better BSR, as this could indicate a more substantial consumer base and higher product credibility.

Furthermore, The Variation Count is a measure of how many different versions or variations of a product are available under the same listing, showing the range of choices offered by a seller. The squared term of variation count, Variation Count Squared, is used to study if there is a point at which offering more variations starts to have a diminishing or negative effect on sales or customer choice. We believe our model will show that a greater variety of options available will positively influence BSR by catering to a broader customer preference spectrum; however, beyond a certain point, too many options may overwhelm consumers and potentially harm BSR.

# of Days Since Last Price Change tracks the time since the product's price was last altered. This can tell us about the seller's pricing strategy—whether they change prices frequently or keep them fixed over time. We predict frequent changes in price will be associated with higher BSR.

Age denotes the time a product has been listed on Amazon, providing insights into its lifecycle—whether it is a new arrival or a long-standing offering. We believe that Age will have a marginal effect on BSR as the ranking is a combination of short and long-term data. Therefore, some products that have been listed on Amazon for less than one year but have a large volume of sales may have a higher BSR than other products that have been listed for multiple years. Conversely, there are situations where products have been listed on Amazon for multiple years and achieve the #1 BSR as they have large amounts of consistent historical data with a high volume of sales.

The # Bought In Past Month measures the quantity of the product sold in the past month, giving an indication of its recent sales performance. This variable could have a large effect on BSR as its an indication of short-term sales performance. We see this variable having a measurable effect on BSR, but, as discussed earlier, the BSR ranking system is proprietary and there could be a smaller weight on the short-term sales for the Vitamin category specifically.

Price represents the cost of a "New" product to the consumer and may influence the product's competitive position in the market. The potential endogeneity of price renders its effect on BSR challenging to isolate.

FBA, or Fulfillment by Amazon, signals whether a seller uses Amazon's fulfillment services, which can affect customer satisfaction and sales through better delivery and customer service. We have seen that participation in FBA will be positively associated with BSR as discussed in the seminal work of Etumnu (2022), as it can enhance delivery efficiency and customer satisfaction. Lastly, Subscribe and Save is a program that offers customers a discount for subscribing to regular deliveries of a product. Participation in this program can be a sign of customer loyalty and can influence repeat purchases leading to an improved BSR.

5.2 Thesis Results: OLS

In our Ordinary Least Squares (OLS) analysis, the coefficients for the variables across six different model specifications provide insights into their impact on the Best Seller Rank (BSR) on Amazon. A negative coefficient indicates an improvement in the BSR as it is ranked from 1 onwards, with 1 being the top sales rank. Table 5 & 6 showcase the coefficients of each variable and we find significance at the 99% confidence level for all variables except Price. An example of an equation is showcased with specification 1 with controls for Vitamin type and month:

$$lnBSR = \beta_0 + \beta_1 SubandSave + \beta_2 FBA + \beta_3 VariationCnt + \beta_4 VariationCntSqd + \beta_5 AvgRating + \beta_6 NumReviews + \beta_7 Age + \beta_8 Price + \beta_9 BoughtInPastMonth + \beta_{10} DaysSincePriceChange$$

The Avg. Rating has coefficients ranging from -0.165 to -0.176, suggesting that as the average rating increases by 1 (on a scale where 1 is low and 5 is high), the BSR improves by about .17%, indicating that products with higher ratings might have better sales. For Number of Reviews (in thousands), the coefficients range from -0.039 to -0.123, indicating a relationship where 1000 additional reviews results in a .039% change in BSR, potentially due to customers having more trust in well-reviewed products as seen in He, Hollenbeck, & Prosperio's (2021) work on fake reviews.

The Variation Count # displayed coefficients between -0.478 and -0.484, indicating that one additional variant count correlates with a better BSR. Sellers can capitalize on this by offering more than one variant on their product listening, potentially to climb the ranks within Amazon's marketplace. This effect is substantial, indicating that deliberate expansion in product variants is a viable strategy to enhance visibility and sales performance on the platform. However, the Variation Count Squared has positive coefficients around 0.03, showing a non-linear pattern with negative returns on each variant past 8 variants. This aligns with Miller's (1956) theory that humans can only prices 7 ± 2 pieces of information at a time.

Also, Days Since Last Price Change showed a small positive coefficient between 0 and 0.008, which could imply that recent price changes are not substantially influencing BSR, however there is a slight reduction in a product's sales rank the longer a seller takes to change the price of their product. However, when we look at the mean of our data set, on average, products make price changes every ~116 days indicating that the trend of infrequent price changes could be to the product's detriment, changing the coefficient to .928. The price has little to no effect on the BSR in our model; this can be attributed to the endogenous nature of price which can be further explored to understand its full effect. Age (in years) of the product listing, with a mean of 7.4 years, has small negative coefficients, around -0.008, implying older products may be at a slight advantage in their BSR. The effect is quite marginal, but this would make sense with Amazon's sales ranking system as the older products have more sales data to contribute to their sales rank.

The Fulfillment by Amazon (FBA) program showed negative coefficients from -0.652 to -0.675, suggesting a robust positive effect on BSR, as products fulfilled by Amazon likely benefit from enhanced customer trust and delivery efficiency. This mirrors the findings of Etumnu (2022) in the coffee industry, who reported that FBA's coefficients were -0.2850 and -0.3256 in different models. Our results extend this understanding by quantifying the impact of FBA in the context of BSR improvement in the Vitamins category, highlighting its effectiveness as a strategic lever for sellers on Amazon. Similarly, Subscribe and Save displayed coefficients from -0.722 to -0.793, indicating that enrollment in this program is associated with significantly better BSR, likely due to customer retention and regular purchases. Lastly, the Number Bought In Past Month with coefficients near -0.0002, however this for each additional sale. The mean number bought in the past month is 261 which indicates that the effect is be a lot larger when dealing with larger volumes of sales as the coefficient would be .0522.

5.3 Alternative Specifications

5.3.1 Lasso Estimation

In the present analysis, the Least Absolute Shrinkage and Selection Operator (LASSO) regression was employed to address the complexities inherent in variable selection and regularization within the dataset. The LASSO extends the capabilities of OLS, effectively shrinking less important variable coefficients to zero. This method is invaluable in our study for two key reasons: it aids in variable selection, identifying the most impactful factors in BSR while mitigating overfitting. This feature is crucial when dealing with high-dimensional data, as is the case with our data set, where numerous variables could potentially influence the outcome. Prior to the application of the LASSO model, the dataset underwent a standardization process using the StandardScaler function from the scikit-learn library, ensuring that each variable was brought to a uniform scale. This preprocessing step creates an effective application of regularization techniques, as it places all variables on equal footing, thereby avoiding undue penalization due to differences in scale.

Following the standardization, the dataset was partitioned into training and testing subsets, with an allocation of 80% for model training and the remaining 20% dedicated to model evaluation. This split was strategically designed to ensure that the model was both trained and validated on representative samples of the data, thereby enhancing the reliability and generalizability of the model's predictive performance.

The LASSO regression model, characterized by an alpha value set to 0.1, was specifically chosen for its propensity to penalize the regression coefficients of less significant variables, effectively reducing them to zero. This penalization serves a dual purpose: it aids in the simplification of the model by eliminating variables with minimal predictive value, and it addresses the issue of multicollinearity by excluding redundant predictors. The regularization effect inherent in the LASSO model is particularly beneficial for datasets with a large number of predictors, as it facilitates the identification of a subset of variables that are most relevant to the dependent variable.

The coefficient estimates derived from the LASSO regression reveal a conservative estimation strategy, described in Table 7, as evidenced by the smaller magnitudes of the coefficients in comparison to those obtained from Ordinary Least Squares (OLS) regression. For instance, the coefficient for "Variation Count" was observed to be -0.149 in the LASSO model, a value notably lower than the range of coefficients estimated by OLS models. This discrepancy brings to light the LASSO model's cautious approach to quantifying the impact of predictors, suggesting a more modest effect of product variety on the Best Sellers Rank (BSR).

Similarly, the coefficients for "Fulfillment by Amazon (FBA)" and the "Subscribe and Save" program, standing at -0.300 and -0.280 respectively, further reflect the LASSO model's propensity towards conservative estimates. These findings, when juxtaposed with the OLS regression results, reveal a deeper understanding of the variables' impacts. Although both variables maintain their significance and direction of effect, their moderated magnitudes within the LASSO regression framework suggest a refined interpretation of how these Amazon services contribute to a product's BSR.

The "Avg. Rating" variable, with a coefficient of -0.184 in the LASSO model, demonstrates a consistency with the OLS findings, reinforcing the understanding of customer ratings' positive relationship with BSR. This consistency across models lends credence to the robustness of the observed effect of product ratings on marketplace success.

Conversely, the "Number of Reviews" presents a smaller coefficient (-0.084) in the LASSO regression than observed in the OLS results, indicating a more conservative estimate of the impact of customer reviews on BSR. This adjustment elucidates the LASSO model's capacity to recalibrate the influence of predictors, offering a refined perspective that underscores the importance of not just the volume of feedback but its qualitative impact.

The variable "Bought In Past Month" exhibits a higher coefficient magnitude (-0.319) in the LASSO model compared to the OLS regressions, suggesting that recent sales performance may wield a more significant influence on BSR than previously estimated. This deviation underscores the LASSO model's unique ability to elevate the importance of short-term sales dynamics, thereby highlighting the critical role of recent market success in enhancing a product's visibility and ranking on Amazon.

In summary, the LASSO regression model, with its emphasis on regularization and variable selection, provides a complementary analytical perspective to that of OLS regression. By imposing a penalty on the regression coefficients, the LASSO model effectively reduces the complexity of the model, thereby facilitating a more nuanced and accurate understanding of the determinants of BSR. This approach not only aids in mitigating the risk of overfitting but also enhances the interpretability of the model by focusing attention on the variables that most significantly influence the dependent variable. The variation in coefficient estimates between the LASSO and OLS models highlights the distinct and valuable contribution of the LASSO regression to econometric analysis, offering deeper insights into the factors that drive BSR within the competitive landscape of the Amazon marketplace.

5.3.2 GLS Regression

In our exploration of the Generalized Least Squares (GLS) regression, as shown in Table 8, we specifically aimed to account for potential deviations from classical regression assumptions, thereby enhancing our analytical robustness. We used the GLS to accommodate the non-normal distributions that characterize our data, such as the skewed nature of review counts and sales ranks. GLSs extend the flexibility of our analysis, allowing us to model data that do not fit the assumptions required by OLS regression. This approach enables us to more accurately capture the effects of new variants, offering a comprehensive view of its impact on Amazon's marketplace.

The GLS model not only corroborates the findings from our Ordinary Least Squares (OLS) analysis but also introduces nuanced variations in effect magnitudes. We specifically use random effects because our dataset may exhibits a hierarchical or clustered structure, with observations grouped in a manner that likely violates the independence assumption of classical regression models. Our data set is split into 4 categories, Vitamin A, B, C, & D so we saw it being beneficial to include random effects as much of the data within each group may have very similarly clustered data points. This nuanced approach is especially evident in our examination of the "Variation Count," where the GLS coefficient (-0.429) provides a pronounced impact on the Best Sellers Rank (BSR). This impact is notably more substantial in comparison to the LASSO model, suggesting that incorporating random effects into our model accentuates the significance of offering a diverse product range on BSR.

The GLS model continues to reveal consistent patterns, such as the "Avg. Rating" displaying a coefficient of -0.170, which indicates a strong and consistent inverse relationship with BSR, mirroring findings from both the OLS and LASSO models. Furthermore, the "Num of Reviews" exhibits a small yet significant coefficient (-0.00007), reinforcing the notion that while reviews do influence BSR, their effect size remains relatively modest. This insight is particularly striking when juxtaposed against OLS results, highlighting the GLS model's capacity to capture finer details that OLS might overlook.

In the case of "Days Since Last Price Change," a positive coefficient (0.0004) remained consistent with the OLS model and confirmed our findings. Similarly, variables like "Bought In Past Month" show comparable effects across both models, emphasizing consistency in findings. The negative coefficients for "Fulfillment by Amazon (FBA)" and "Subscribe & Save" (-0.101 and -0.179, respectively) further reinforce the advantage of Amazon's fulfillment services in enhancing BSR. This finding aligns with OLS results but with moderated magnitudes in GLS, hinting at a more thorough understanding of FBA's positive impact after factoring in random effects.

This exploration emphasizes the crucial role of accounting for random effects in GLS regression, which significantly enriches our understanding of each variable's true influence on BSR. By adjusting for these effects, the GLS model provides a more detailed and accurate representation of variable relationships, offering a deeper layer of analytical insight beyond the initial observations from OLS and LASSO models.

Our multi-faceted econometric approach, utilizing OLS, LASSO, and GLS regressions, reveals the intricate dynamics at play in determining BSR. While the OLS models provide a broad understanding of the variables' impacts, the Lasso regression refines this understanding by selecting the most impactful predictors. The consistent significance of certain variables across all models underscores the reliability of our analytical framework and the validity of our findings in the context of Amazon's marketplace.

5.4 Limitations

This thesis encountered several limitations that prevented the research from further developing. First, our analysis lacked access to search order ranking data, which could significantly influence product visibility and sales on Amazon, greatly affecting BSR. The absence of this data means our findings primarily reflect correlations with Amazon's Best Seller Ranking without fully capturing the dynamics of product discoverability through search results.

Secondly, the scope of our dataset stayed confined to six months. While this timeframe allowed us to observe short-term trends and effects, keeping products fixed it may not adequately capture long-term strategic shifts among sellers or the full impact of Amazon's algorithm changes over time. Additionally, having long-term data would be difficult to keep products fixed over that period due to the granular and quickly changing BSR market. Seasonal variations, product life cycles, and market entry or exit of contenders within this limited period could also skew the findings.

Another limitation is our study's focus on observable metrics such as product variants, fulfillment method, and participation in Amazon's Subscribe and Save program. This focus omits potentially influential but unmeasured factors like marketing strategies, seller reputation, and external traffic to product listings, which could also affect BSR. While our analysis suggests a strategic advantage in using Amazon's FBA service and the Subscribe and Save feature, we still need to explore the cost implications of these strategies for sellers fully. The financial burden of FBA fees and discounts for subscription services might offset or significantly impact the net benefits of improved BSR.

Furthermore, the analysis encounters a critical limitation in its reliance on a broad dataset for estimating the effects on Amazon's Best Seller Ranking (BSR), rather than concentrating solely on data specific to Bountiful. This approach inherently dilutes the specificity of our findings to Bountiful's unique market situation. Particularly, while data on practices such as review hijacking pertinent to Bountiful were accessible, the limited number of data points specific to Bountiful constrains the robustness and predictive accuracy of our model. This scarcity of Bountiful-specific data points not only restricts our ability to construct a compelling model uniquely tailored to Bountiful's market dynamics but also potentially overlooks the nuanced effects of such practices on Bountiful's BSR and overall market performance. Consequently, while the analysis offers insights into general trends affecting BSR in the vitamin category, the extrapolation of these findings to Bountiful's case requires cautious interpretation, given the model's limited capacity to fully encapsulate the intricate effects of targeted strategies and market behaviors unique to Bountiful.

5.5 Discussion

The findings of this thesis illuminate the intricate influence of product variants on Amazon's Best Seller Ranking within the Vitamins and Supplements sector. While variants can enhance consumer choice by catering to diverse preferences, our analysis reveals that they can also be wielded as a strategic tool to artificially boost sales rankings, potentially misleading consumers.

To forge a path toward effective regulation of variant creation, it is critical to weigh the costs and benefits from multiple vantage points. Sellers may be incentivized to multiply variants to capitalize on increased visibility and sales, but this strategy may not always be aligned with buyer interests. Consumers may initially benefit from the personalized options and product diversity that variants provide, enabling a more tailored shopping experience. However, this benefit has a threshold; beyond a certain point, additional variants may result in decision paralysis and quality assessment challenges. For the platform, while a broad selection enhances the richness of offerings, the resulting complexity can dampen the shopping experience and pose significant quality control hurdles.

The interplay of these factors suggests that while variant creation offers advantages, its unchecked growth could introduce inefficiencies and distrust into the marketplace. To curb potential exploitation, this thesis proposes strategic interventions, including the novel designation of new variants as "New," serving as a clear signal to consumers navigating their options.

To bolster real-time surveillance, a dedicated analysis team, augmented by a sophisticated machine learning model, could detect and preclude review manipulation attempts. These proposed solutions aim to advance the marketplace's preemptive defenses, reinforcing the authenticity of sales ranks and preserving consumer trust.

Moving forward, comprehensive research should include a wider array of variables and a more extended observation window to discern enduring trends. The interplay between Amazon's marketplace strategies and their influence on sales rank, especially over several years, requires further exploration.

Furthermore, integrating search rank data can offer insights into how product visibility is shaped by variant strategies and review practices. Such an understanding is crucial for cultivating a fair marketplace, where consumer choices drive success rather than being swayed by manipulative tactics.

This thesis ventures beyond identifying current market intricacies; it proposes actionable solutions that could reshape the e-commerce landscape, emphasizing transparency and fairness. By advancing this conversation, we contribute to the ethical discourse surrounding digital market strategies and endorse the creation of a digital marketplace that not only serves the economic interests of sellers and the platform but also upholds the integrity of consumer choice. The balance of these interests is pivotal, as it supports a sustainable model where the marketplace is a trusted environment, fostering fair competition and providing authentic product differentiation.

In sum, the actionable insights offered by this study call for thoughtful policy reforms and platform design changes that align closely with the core values of market fairness and consumer protection. As we consider the future of e-commerce, it is these values that will secure the long-term health and vibrancy of digital marketplaces, ensuring they continue to serve as engines of innovation and choice in the global economy.

6 Conclusion

The investigation into The Bountiful Company's review hijacking on Amazon, as detailed in this thesis, brings to light the intricacies and challenges of maintaining honest consumer feedback in an era dominated by e-commerce. By exploiting Amazon's system of variation listings, The Bountiful Company was able to artificially inflate their product visibility and sales, showcasing a glaring loophole in digital market regulation. This case study not only exposes a critical vulnerability in online retail but also prompts a necessary discussion on the implementation of more robust mechanisms to protect the integrity of consumer reviews.

Moreover, the empirical findings of this research offer a foundation for developing strategies aimed at preventing review hijacking. The proposed measures, including the identification of new product variants and enhanced monitoring for review manipulation, suggest a path forward in reconciling the need for marketplace innovation with the imperative of consumer protection. In essence, the conclusions drawn from this study advocate for a more vigilant and transparent online shopping environment, where consumer trust is not undermined by deceptive marketing tactics.

References

[1] "Amazon Review Hijacking – How to Spot Sellers That Are Recycling Reviews on Amazon - ReviewMeta Blog." Accessed March 7, 2024. https://ReviewMeta.com/blog/amazon-review-hijacking/.

[2] "Best Seller Rank (BSR) to Sales: An Empirical Look at Amazon.Com | IEEE Conference Publication | IEEE Xplore." Accessed November 6, 2023. https://ieeexplore.ieee.org/abstract/document/9282620.

[3] "Brand Name Types and Consumer Demand: Evidence from China's Automobile Mark...: EBSCOhost." Accessed January 5, 2024. https://web-p-ebscohost-com.proxy.library.emory.edu/ehost/pdfviewer/pdfviewer?vid=1&sid=c76af0a1-46e5-404a-ba78-389f9e65b19b%40redis.

[4] Chevalier, Judith A., and Dina Mayzlin. "The Effect of Word of Mouth on Sales: Online Book Reviews." Journal of Marketing Research 43, no. 3 (2006): 345–54.

[5] "Creating and Detecting Fake Reviews of Online Products - ScienceDirect." Accessed November 3, 2023. https://www.sciencedirect.com/science/article/pii/S0969698921003374.

[6] "Emotional Expressions and Brand Status." Accessed January 5, 2024. https://doi.org/10.1177/00222437211037340.

[7] Etumnu, Chinonso E. "A Competitive Marketplace or an Unfair Competitor? An Analysis of Amazon and Its Best Sellers Ranks." Journal of Agricultural Economics 73, no. 3 (September 2022): 924–37.

[8] Federal Trade Commission. "Advertising and Marketing on the Internet: Rules of the Road," December 12, 2000. https://www.ftc.gov/business-guidance/resources/advertising-marketing-internet-rules-road.

[9] Federal Trade Commission. “Consumer Review Fairness Act: What Businesses Need to Know,” February 21, 2017. https://www.ftc.gov/business-guidance/resources/consumer-review-fairness-act-what-businesses-need-know.

[10] Federal Trade Commission. “Federal Trade Commission Announces Proposed Rule Banning Fake Reviews and Testimonials,” June 29, 2023. https://www.ftc.gov/news-events/news/press-releases/2023/06/federal-trade-commission-announces-proposed-rule-banning-fake-reviews-testimonials.

[11] Federal Trade Commission. “FTC’s Endorsement Guides: What People Are Asking,” September 7, 2017. https://www.ftc.gov/business-guidance/resources/ftcs-endorsement-guides-what-people-are-asking.

[12] Federal Trade Commission. "In the Matter of The Bountiful Company, a corporation." Docket No. C-4791. March 28, 2023.

[13] Flisikowski, Karol, and Wioleta Kucharska. “The Condition of Economies. Do Most Valuable Global Brands Matter?” Equilibrium 13, no. 2 (June 2018): 251–64. https://doi.org/10.24136/eq.2018.013.

[14] Goldstein, Eve. 2010. "Price Dispersion and Pricing Strategies for Firms on Amazon.com." Bachelor’s thesis, Emory College of Arts and Sciences, Emory University.

[15] He, Sherry, Brett Hollenbeck, and Davide Proserpio. “The Market for Fake Reviews.” SSRN Scholarly Paper. Rochester, NY, October 1, 2022. https://doi.org/10.2139/ssrn.3664992.

[16] “Here Are Some Ways To Tell If Amazon Reviews Are Fake.” Accessed March 7, 2024. https://www.buzzfeednews.com/article/nicolenguyen/fake-amazon-reviews-black-friday-cyber-monday-deals.

[17] Karabas, Ismail, Ioannis Kareklas, Tj Weber, and Darrel Muehling. "The Impact of Review Valence and Awareness of Deceptive Practices on Consumers' Responses to Online Product Ratings and Reviews." Journal of Marketing Communications 27 (May 5, 2020): 1–31. https://doi.org/10.1080/13527266.2020.1759120.

[18] Lee, Jung-Yong, and Chang-Hyun Jin. "The Role of Ethical Marketing Issues in Consumer-Brand Relationship." Sustainability 11, no. 23 (January 2019): 6536. https://doi.org/10.3390/su11236536.

[19] Luca, Michael, and Georgios Zervas. "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud." SSRN Scholarly Paper. Rochester, NY, May 1, 2015. https://doi.org/10.2139/ssrn.2293164.

[20] Mayzlin, Dina, Yaniv Dover, and Judith Chevalier. "Promotional Reviews: An Empirical Investigation of Online Review Manipulation." American Economic Review 104, no. 8 (August 2014): 2421–55. https://doi.org/10.1257/aer.104.8.2421.

[21] Moon, Sangkil, Moon-Yong Kim, and Dawn Iacobucci. "Content Analysis of Fake Consumer Reviews by Survey-Based Text Categorization." International Journal of Research in Marketing 38, no. 2 (June 1, 2021): 343–64. https://doi.org/10.1016/j.ijresmar.2020.08.001.

[22] Salmen, Alexander. "New Product Launch Success: A Literature Review." Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis 69 (March 1, 2021): 151–76. https://doi.org/10.11118/actaun.2021.008.

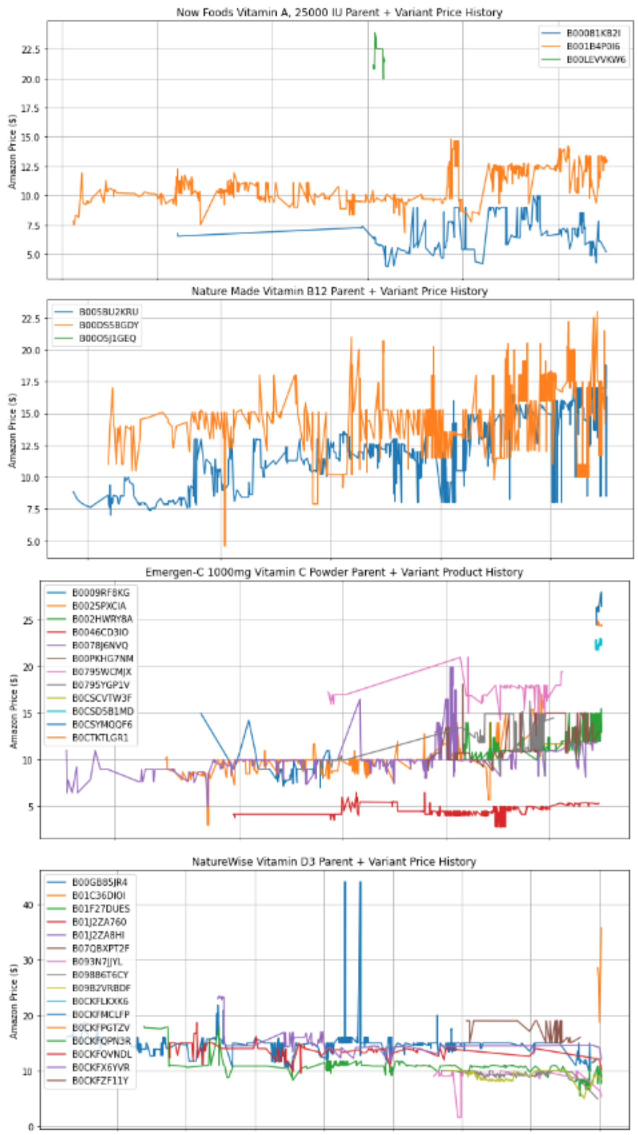
[23] Sorokina, Daria, and Erick Cantu-Paz. "Amazon Search: The Joy of Ranking Products." In Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval, 459–60. Pisa Italy: ACM, 2016. https://doi.org/10.1145/2911451.2926725.

[24] Skrovan, Sandy. "The Impact of Amazon Badges on Traffic and Conversion." Accessed March 28, 2024. https://www.profitero.com/blog/2018/11/the-impact-of-amazon-badges-on-traffic-and-conversion.

[25] Swearingen, Jake. "Hijacked Reviews on Amazon Can Trick Shoppers." Consumer Reports, August 26, 2019. https://www.consumerreports.org/customer-reviews-ratings/hijacked-reviews-on-amazon-can-trick-shoppers/.

Appendix

Figure 3: Line Graphs of Price Data for Vitamins A, B, C, & D



Note: This graph illustrates the price fluctuations of the #1 Best Sellers for Vitamins A, B, C, and D. It showcases the variations in pricing among different product variants over each product's life span. The depicted trends reflect the dynamic nature of the market and consumer preferences that influence the pricing strategies for these top-selling vitamin products.

Table 1: Keepa Variable Descriptions

Variable	Description
Sales Rank Current	The current sales rank of the product.
Bought in Past Month	Quantity of the product bought in the past month.
Reviews - Rating	The average customer review rating.
Reviews - Review Count	The total number of customer reviews.
Ratings - Format Specific	Ratings specific to the product format.
Review Count - Format Specific	Review count specific to the product format.
Buy Box: Is FBA	Indicates if the Buy Box is fulfilled by Amazon (FBA).
New, Current (Price in $)	The current price for new condition.
ASIN	Amazon Standard Identification Number, a unique identifier for the product.
Age	The # of years a product has been listed on Amazon.com
Subscribe & Save	The Product Listing contains the Subscribe & Save feature
Days Since Price Change	# of days since a product has undergone a price change on Amazon.com
Variation ASIN Count	The # of variants a product has on its Amazon Listing Page

Table 2: Variable Categories

Categorical Columns	**Numeric Columns**	**Binary Columns**	**Time Columns**
Category	Sales Rank: Current	Buy Box: Is FBA	Last Price Change
Variation ASINs	Variation ASIN Count	Subscribe and Save	Tracking since
ASIN	Bought in past month		Age
Variation Attributes	Reviews: Review Count		
	New: Current		

Table 3: Regression Variables Summary Statistics

Variable	count	mean	std	min	25%	50%	75%	max
lnBSR	57618	12.188	1.288	2.708	11.497	12.488	13.127	14.005
Avg Rating	57618	4.066	1.367	0.0	4.2	4.5	4.7	5.0
Num Reviews (in 1000s)	57618	0.556	3.191	0.0	0.007	0.041	0.208	161.76
Variation Cnt	57618	1.372	1.020	1.0	1.0	1.0	1.0	27.0
Variation Cnt Sqrd	57618	2.923	10.551	1.0	1.0	1.0	1.0	729.0
Days Since Price Change	57618	115.987	168.799	17.0	21.0	41.0	136.0	2670.0
Age	56774	7.381	5.769	0.082	2.5178	5.836	11.151	25.219
Bought In Past Month	57618	261.023	1566.399	0.0	0.0	0.0	50.0	100000.0
Price	57618	20.207	41.790	0.0	6.99	16.47	26.06	5299.99

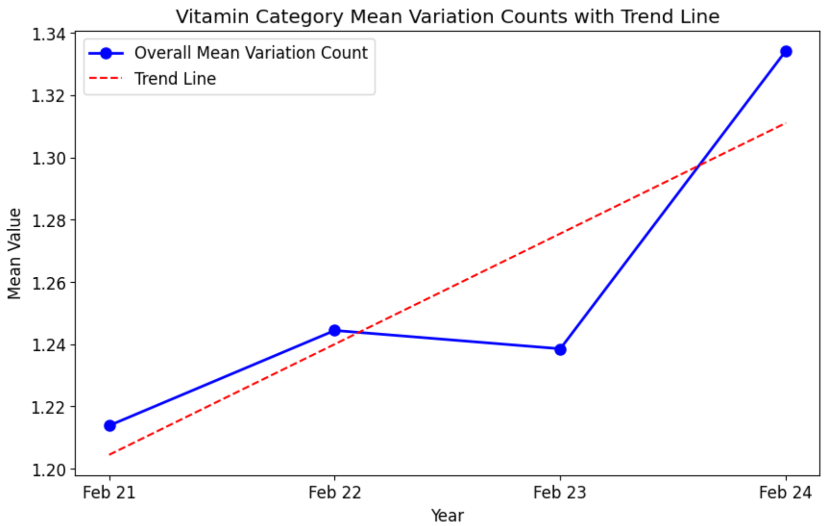
Sub and Save	57618	0.238	0.426	0.0	0.0	0.0	0.0	1.0
FBA	57618	0.513	0.450	0.0	0.0	1.0	1.0	1.0

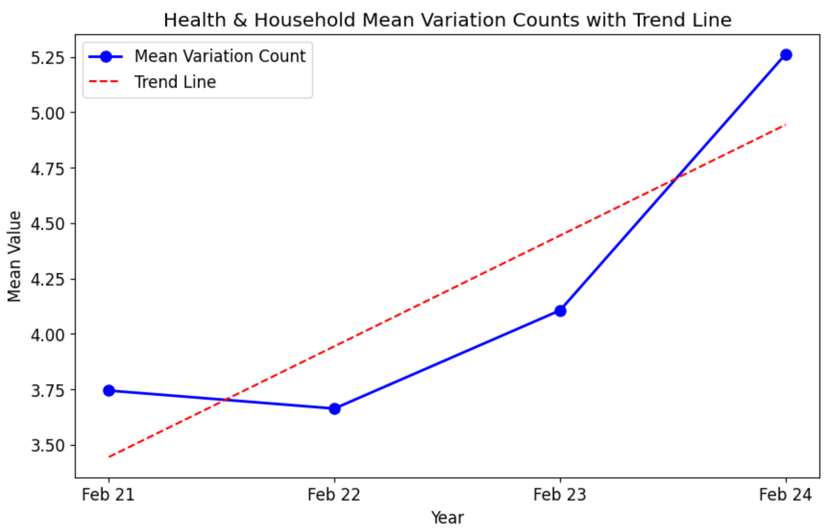
Table 4: Correlation Matrix

	lnBSR	Avg Rating	Num Reviews	Variation Cnt	Variation Cnt Sqrd	Age	Bought In Past Month	FBA	Price	Sub and Save
lnBSR	1.0	-0.338	-0.43	-0.319	-0.184	-0.076	-0.484	**-0.524***	-0.01	**-0.529***
Avg Rating		1.0	0.071	0.066	0.016	0.188	0.068	0.168	-0.061	0.16
Num Reviews			1.0	0.238	0.224	0.066	0.764	0.11	-0.014	0.149
Variation Cnt				1.0	**0.835****	0.01	0.175	0.095	0.071	0.144
Variation Cnt Sqrd					1.0	-0.005	0.176	0.043	0.063	0.079
Age						1.0	0.044	-0.063	-0.015	-0.053
Bought In Past Month							1.0	0.145	-0.009	0.201
FBA								1.0	0.082	**0.532***
Price									1.0	0.036
Sub and Save										1.0

Note: ** indicates that the variables and highly correlated ($|x| > 0.7$) and * indicates that the variables are moderately correlated ($0.5 \leq |x| \leq 0.7$)

Figure 4: Variation ASIN Count Trend Lines by Product & Vitamin Categories





T-statistics with P-values for Vitamin pairwise t-test:

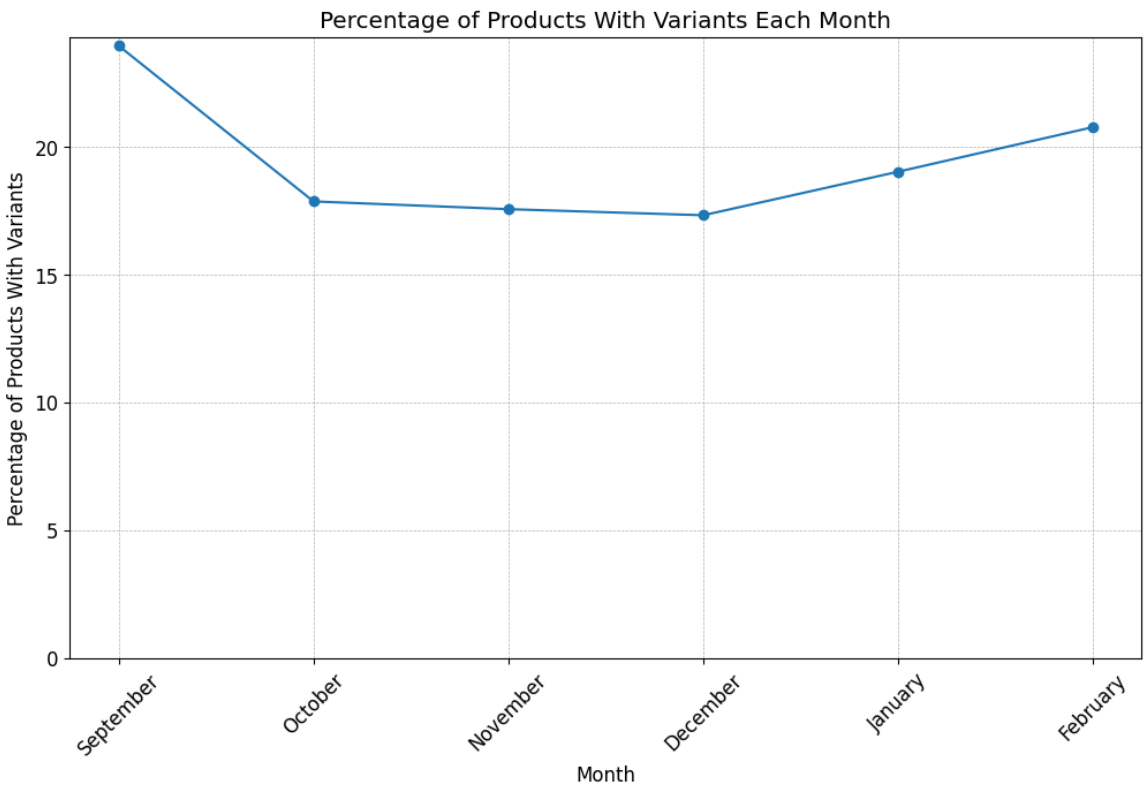
	Year1	Year2	T-Statistic	P-Value
0	2021	2022	1.580259	1.140611e-01
1	2021	2023	−0.707178	4.794628e-01
2	2021	2024	−26.074633	1.696981e-147
3	2022	2023	−2.311273	2.082631e-02
4	2022	2024	−26.924569	5.919014e-157
5	2023	2024	−26.152144	2.346533e-148

T-statistics with P-values for Health & Household pairwise t-test:

	Year1	Year2	T-Statistic	P-Value
0	2021	2022	0.387615	6.983094e-01
1	2021	2023	-1.384283	1.663056e-01
2	2021	2024	-6.120550	9.695060e-10
3	2022	2023	-1.735646	8.266167e-02
4	2022	2024	-6.620797	3.773283e-11
5	2023	2024	-4.018109	5.909996e-05

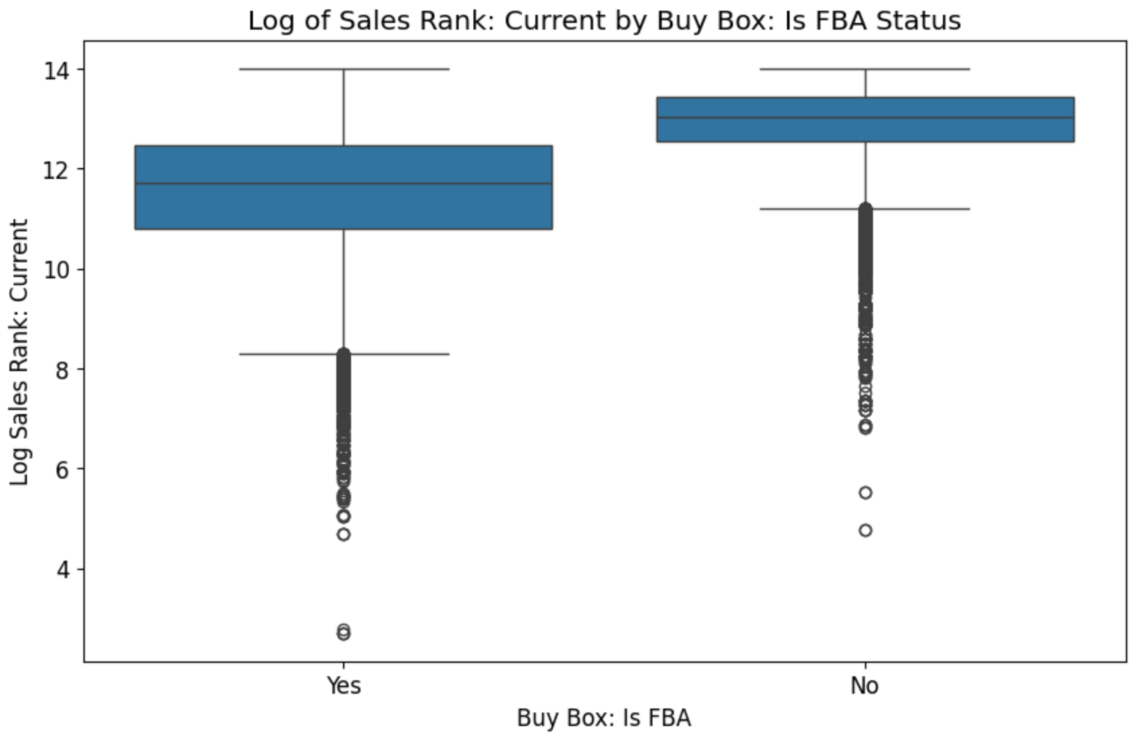
Note: We find a statistically significant difference between the variant counts in 2021 and 2024 for both the Vitamin category and the Health & Household.

Figure 5: Line Graph Representing Percentage of Products With Variants



Note: This graph is showcasing the percentage of products with variants over the span of 6 months in our data set. The data points hover between 17% and 24% pointing to a possible trend of the proportion of products with variants being the minority in each month's data sets.

Figure 6: Box-plot of Log(Sales Rank: Current) Divided by FBA



Note: Mann-Whitney U Test results -- U-statistic: 141465068.0, P-value: 0.0 & T-test results -- T-statistic: -149.600, P-value: 0.0

Table 5: OLS Specification Outputs

	$\hat{\beta}_{M1}$	$s(\hat{\beta}_{M1})$	$\hat{\beta}_{M2}$	$s(\hat{\beta}_{M2})$	$\hat{\beta}_{M3}$	$s(\hat{\beta}_{M3})$	$\hat{\beta}_{M4}$	$s(\hat{\beta}_{M4})$	$\hat{\beta}_{M5}$	$s(\hat{\beta}_{M5})$	$\hat{\beta}_{M6}$	$s(\hat{\beta}_{M6})$
Intercept	13.999	0.042	14.032	0.034	14.012	0.041	14.049	0.032	13.975	0.036	13.949	0.035
AvgRating (1-5)	-0.165	0.042	-0.174	0.002	-0.166	0.002	-0.176	0.002	-0.164	0.002	-0.174	0.002
NumReviews (in thousands)	-0.039	0.005	-0.038	0.005	-0.039	0.005	-0.039	0.005	-0.123	0.007	-0.123	0.007
VariationCnt	-0.478	0.032	-0.485	0.032	-0.477	0.032	-0.484	0.032	-0.453	0.023	-0.451	0.023
VariationCntSqrd	0.031	0.005	0.032	0.005	0.032	0.005	0.032	0.005	0.029	0.003	0.029	0.003
DaysSincePriceChange	0.008	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000
Age (in Years)	-0.008	0.001	-0.007	0.001	-0.008	0.001	-0.007	0.001	-0.008	0.001	-0.008	0.001
FBA	-0.652	0.009	-0.675	0.009	-0.648	0.009	-0.670	0.009	-0.667	0.009	-0.675	0.009
SubandSave	-0.722	0.018	-0.730	0.018	-0.723	0.018	-0.731	0.018	-0.790	0.012	-0.793	0.012
Price (in dollars)	-0.001	0.000	0.001	0.000					0.001	0.000	0.001	0.000
BoughtInPastMonth	-0.0002	0.000	-0.0002	0.000	-0.0002	0.000	-0.0002	0.000				
Month	Yes		Yes		Yes		Yes		Yes		No	
Category	Yes		No		Yes		No		Yes		Yes	
σ	62.1		63.0		62.0		61.3		58.9		58.7	
Sample Size	56774		56774		56774		56774		56774		56774	

Note: The LHS variable of these regressions is the ln(BSR). All variables are significant at the 99% confidence level except Price. $\hat{\beta}$ is defined as the coefficient of each variable for the 6 specifications and s() is the standard error of each variable for the respective specification. σ is defined as the R^2 of each specification. We determined the negative effects of adding additional variants by taking the derivative of specification 1 then finding the number of variants where each additional ultimately leads to negative returns.

Table 6: Linear Regression Output Specifications

	VariationCnt	VariationCntSqrd	FBA	SubandSave	$\widehat{\sigma}$	Sample Size
$\widehat{\beta}_{M1}$	-0.478	0.031	-0.652	-0.722	62.1	56774
$s(\widehat{\beta})_{M1}$	0.032	0.005	0.009	0.018		
$\widehat{\beta}_{M2}$	-0.485	0.032	-0.675	-0.730	63.0	56774
$s(\widehat{\beta})_{M2}$	0.032	0.005	0.009	0.018		
$\widehat{\beta}_{M3}$	-0.477	0.032	-0.648	-0.723	62.0	56774
$s(\widehat{\beta})_{M3}$	0.032	0.005	0.009	0.018		
$\widehat{\beta}_{M4}$	-0.484	0.032	-0.670	-0.731	61.3	56774
$s(\widehat{\beta})_{M4}$	0.032	0.005	0.009	0.018		
$\widehat{\beta}_{M5}$	-0.453	0.029	-0.667	-0.790	58.9	56774
$s(\widehat{\beta})_{M5}$	0.023	0.003	0.009	0.012		
$\widehat{\beta}_{M6}$	-0.451	0.029	-0.675	-0.793	58.7	56774
$s(\widehat{\beta})_{M6}$	0.023	0.003	0.009	0.012		

Note: $\widehat{\sigma}$ is the R^2 of each specification

Table 7: Lasso Estimation Output

Variables	**Coefficient**
DaysSincePriceChange	0.076
Category_Vitamin C	0.025
Category_Vitamin B	-0.016
NumReviews	-0.084
VariationCnt	-0.149
AvgRating	-0.184
SubandSave	-0.280
FBA	-0.300
BoughtInPastMonth	-0.319

Note: The data is standardized using the StandardScaler from scikit-learn, ensuring that all variables contributed equally to the analysis by giving them a common scale. With the preprocessed data, we split our dataset, allocating 80% to training and 20% to testing, ensuring a representative sample for model training and evaluation. We choose the Lasso regression model with an alpha value of 0.1 for its ability to perform variable selection, penalizing less significant variables to zero.

Table 8: GLS Regression Output

Variables	**Coefficient**
Price	0.000
Avg. Rating	-0.170***
NumReviews	-0.000***
VariationCnt	-0.429***
VariationCntSqrd	0.216***
DaysSincePriceChange	0.000***
BoughtInPastMonth	-0.000***
FBA	-0.101***
SubandSave	-0.179***

Note: All variables except Price are significant at the 99% confidence level ($p < 0.01$).