Sustainability and Competition on Amazon[1]

Xiaohang (Flora) Feng, Carnegie Mellon University, xiaohanf@andrew.cmu.edu;
Xiao Liu, New York University, xl23@stern.nyu.edu;
Shunyuan Zhang, Harvard University, szhang@hbs.edu;
Kannan Srinivasan, Carnegie Mellon University, kannans@cmu.edu.

Abstract

Amazon introduced the Climate Pledge Friendly (CPF) badge by consolidating various green certificates to examine its impact on market dynamics. We applied a game-theoretic model and causal inference using data from Amazon.com to explore the effects of this badge on consumer behavior, seller pricing, and market concentration. Our theoretical model outlines a three-stage process where sellers set prices, the marketplace determines badge eligibility, and consumers make purchase decisions. We discovered that increased demand, higher prices, and reduced market concentration occur when the benefits gained from attracting green consumers exceed the detriments from alienating non-green consumers due to increased prices. Optimal conditions were identified where certifying only the most sustainable products maximizes outcomes over strategies that result in either all or no products being badged. Empirically, we gathered six months of data on 6,606 products across eight categories, using the interactive fixed effect counterfactual (IFEct) estimator to manage endogeneity and treatment reversals. Our findings indicate that the CPF badge significantly enhances sales volume, increases product prices, and decreases market concentration. These results guide sellers considering green certification and platforms contemplating unified green badge policies.

[1] The authors thank the participants at the Marketing for Environmental Sustainability Conference and the Theory and Practice in Marketing Conference for their valuable feedback and suggestions. In particular, we would like to thank Raphael Thomadsen and Bryan Bollinger for their insightful comments.

Keywords: Sustainability, Eco-labeling, E-commerce, Seller Competition, Causal Inference, Multimodal Vector Representation

1. Introduction

Sustainability and green marketing are significant trends, with sustainable product sales expected to grow by over 19.5% from 2023 to 2032 (Global Market Insight). According to Forbes, in 2023, 90% of business leaders view sustainability as crucial, and 60% of companies have developed sustainability strategies. As eco-consciousness rises among consumers and businesses, marketplaces are incentivized to offer sustainability certifications. In 2020, reflecting its commitment to sustainability, Amazon introduced the Climate Pledge Friendly (CPF) badge, and as of September 2023, has collaborated with 50 external certifiers and established two of its own certifications.[2] Any product sold with one of these 52 certifications may have the CPF badge; thus, it is a unifying green badge representative of all these 52 certifications. We define such a policy as a unified green badge policy.

The sales impact of a green badge is complex. On the one hand, raising eco-consciousness among consumers may boost sales, especially for brands that can compete more effectively in sustainability. On the other hand, a strong focus on sustainability might detract from product design, reducing its appeal and sales potential (White, Hardisty, and Habib, 2019; Townsend and Shu, 2010). As a result, the badge could hurt sales if consumers see the products as overpriced, of lower quality, or poorly designed. That is, consumers may be reluctant to pay a premium for sustainable products, and sustainable materials might compromise functionality, which can be viewed as a drawback (Luchs et al., 2010; Newman, Gorlin, and Dhar, 2014).

It follows that the badge's influence on pricing is also twofold: it could lead to higher prices if consumers are willing to pay more for green products, but it might lower prices if there is a gap between stated willingness to pay and actual purchasing behavior (Talwar et al., 2021). Furthermore, the effect of a green badge on market concentration remains underexplored, with limited research suggesting that

[2] See a detailed introduction at https://www.amazon.com/b?node=21221608011.

tailoring products for different consumer segments could reduce competitors' profits (Amaldoss and Prusty, 2024).

Given the complex effects of green badging on sales and pricing and the limited research on how they affect market concentration, this paper seeks to address this gap in the literature. By integrating insights from a three-stage game theory model with empirical evidence from Amazon sales data, we offer the mechanisms and necessary conditions that lead to increased demand, higher prices, and reduced market concentration following the adoption of a green badge. The empirical analysis corroborates the theoretical model, confirming significant benefits derived from a unified green badge: increased demand benefits both the sellers and the marketplace, higher prices favor the sellers and the marketplace, and reduced market concentration benefits both the marketplace and small sellers.

We first develop a theoretical game theory model and document a specific condition that facilitates a scenario where the adoption of green badges leads to increased price and demand, coupled with reduced market concentration. This balance is achieved when the increase in demand from eco-conscious consumers compensates for any decrease due to higher prices that deter non-green consumers from purchasing. Specifically, the appeal of sustainable features may enhance demand among green consumers, yet a price premium on badged products could dissuade less environmentally focused customers from buying those products. Consequently, the intrinsic appeal of a product's non-sustainability-related attributes must be robust enough to ensure that the gains from green consumers outweigh the losses from those deterred by higher prices. Our findings further suggest that an optimal balance occurs when the badging proportion is around 50%. Under these conditions, both the marketplace and the sellers can maximize their profits compared to scenarios where the badging proportion is either 0% or 100%. This optimal scenario arises when the baseline utility for the non-sustainability-related features of one product slightly surpasses that of the other, and the baseline utility for sustainability-related features remains substantially lower than that for non-sustainability attributes of both products.

Next, we offer empirical evidence of the multifaceted impacts of adopting a green badge in online marketplaces. We collected daily data spanning from March 1, 2023, to September 15, 2023, resulting in

6,606 unique products across eight product categories. In our data, 35.39% of the products are badged, closely approximating a badge threshold of 50% as predicted by our theoretical model. Thus we conjecture that the adoption of the CPF badge is associated with increased demand and price, as well as reduced market concentration.

To causally examine the impact of the badge on market outcomes, we used the interactive fixed effect counterfactual (IFEct) estimator (Liu, Wang, and Xu 2022). This approach is particularly designed to handle treatment reversal scenarios—where the treatment applied to a group is subsequently reversed or altered—and to address endogeneity concerns associated with unobserved counterfactuals through an augmented factor structure (Bai 2009). Treatment reversal exists in our dataset, because 29.79% of the products first received a CPF badge, then lost it at a later period (some might subsequently regain it). Further, to control for potential confounding variables, we included various time-varying covariates extracted from product reviews, descriptions, and images. Additionally, we employed machine learning models to derive insights from unstructured data sources such as product images and reviews, enhancing the robustness of our estimates.

Our study presents three key findings. First, the CPF badge significantly increases overall product demand, particularly for products targeting older individuals and males, compared to those aimed at younger and female demographics. Second, the badge is associated with a significant price increase, consistent with the literature on sustainable goods carrying a price premium (Tully and Winer 2014), and this price effect holds across most categories. Last, the CPF badge enhances the competitiveness of smaller brands, leading to reduced market concentration within subcategories, suggesting that the badge fosters a more competitive environment for small brands on the platform. To ensure robustness, we used multimodal machine learning models to derive vector representations of images and descriptions, addressing concerns about feature selection and validating our findings.

Our research represents a pioneering effort to examine the profound effects of a unified eco-label on demand, price, and market concentration across an e-commerce platform. We elucidate the mechanisms through which the adoption of such a unified green badge augments demand, elevates price levels, and reduces market concentration. This study not only clarifies the direct benefits of eco-labeling but also highlights its role in fostering a more competitive and sustainable marketplace.

2. Literature Review

We contribute to three strands of literature: the drivers of sustainability, the motivations behind brand sustainability marketing, and the impact of eco-labeling on product demand and pricing.

2.1 Drivers of Sustainable Consumption

Previous research on drivers of sustainable consumption primarily focuses on two areas: product attributes and consumer features.

Product attributes. Research in this field often examines the nuanced impact of green attributes on product evaluations. Luchs et al. (2010) found that green products are often associated with attributes such as gentleness rather than strength, leading consumers to prefer conventional products when strength is a priority for them. Chen and Chang (2013) also noted that consumers frequently perceive green products as less effective than their traditional counterparts. Additionally, Newman, Gorlin, and Dhar (2014) showed that when companies prominently highlight green attributes, consumers may assume that resources are being diverted from improving functionality, reducing perceptions of quality and purchase intentions. These studies highlight how the framing of environmental efforts can significantly influence consumer perceptions and preferences.

Consumer features. Sustainable consumption behaviors are significantly shaped by demographic factors including age, gender, and education (Murphy, Kangun, and Locander 1978). Women generally display higher levels of sustainable behavior, possibly due to traits such as agreeableness and openness (Luchs and Mooradian 2012). Green consumers are often seen as more cooperative and ethical (Mazar and Zhong 2010), while younger, more liberal, and highly educated individuals are more likely to engage in pro-environmental actions (Luchs and Mooradian 2012). Additionally, intelligence, education, and knowledge positively correlate with greater responsiveness to environmental appeals (Aspara, Luo, and Dhar 2017).

No prior studies have examined how different demographic groups respond to a unified green badge on Amazon. Our findings diverge from existing research on the correlation between demographics and interest in sustainable products. We find that the CPF badge significantly boosts demand for products preferred by men, with no significant effect on products favored by women. Additionally, we find the badge's positive demand effect is stronger for products preferred by older individuals. A unified badge increases demand among those with limited knowledge of sustainable goods, as opposed to multiple green certificates, which can cause information overload and make it harder for less sustainability-focused consumers to differentiate between the two categories of products.

2.2 Motivations for Sustainability Marketing

Existing literature highlights three principal motivations for businesses to engage in sustainability marketing. Firstly, companies that adapt to the evolving landscape, particularly by addressing the need for sustainability, position themselves for long-term survival and prosperity, gaining strategic advantages (Banerjee, Iyer, and Kashyap 2003). Secondly, studies consistently show that adopting socially and environmentally responsible practices improves consumer perceptions and leads to greater profitability (Olsen, Slotegraaf, and Chandukala 2014; Sen and Bhattacharya 2001). Thirdly, firms that embrace sustainable operations and pioneer innovative business models for sustainable consumption often achieve higher long-term profits, as demonstrated by the successes within the sharing economy. While traditional marketing efforts have focused on identifying the green consumer, modern research emphasizes understanding the predictors of sustainable consumption (Menon and Menon 1997). This shift prompts marketers to expand their strategies, offering mutual long-term benefits to both businesses and the environment. As companies pursue more environmentally sustainable practices, it becomes crucial to cultivate recognition and rewards from consumers for the companies' sustainable values, potentially fostering sustainable consumption and enhancing the companies' sustainability and strategic positioning.

Our contribution to this body of research centers on examining firms' potential profitability through strategic price adjustments and modifying product images and descriptions on an e-commerce platform, framed within a unified green badge policy. This area of inquiry is novel, as prior research has yet to delve into such a scenario, likely due to the recent implementation of unified green certification practices by e-commerce platforms such as Amazon.

2.3 The Impact of Eco-Labeling

Previous research has focused on how eco-labeling affects demand and pricing. Eco-labels communicate a product's sustainable attributes, helping consumers make informed eco-friendly choices (Parguel, Benoît-Moreau, and Larceneux 2011). However, information overload, limited exposure, and confusion can hinder sustainable behavior (Chen and Chang 2013). Influential labels that are attention-grabbing, easy to understand, and consistent across categories can better guide eco-friendly decisions. Contrasting positive labels with negative ones that highlight harmful attributes may further improve label effectiveness (Borin, Cerf, and Krishnan 2011).

The pricing of eco-labeling depends on consumers' willingness to pay a premium for sustainable products (Tully and Winer 2014). Yet, a gap often exists between stated willingness to pay and actual purchase behavior (Johnstone and Tan 2015). Third-party certification can enhance the transparency and credibility of eco-labels. Still, the proliferation of various green certificates may undermine their effectiveness due to concerns about credibility and certifying bodies (Borin, Cerf, and Krishnan 2011).

Our study is among the first to explore the effects of a unified green badge—a consolidation of various green certificates—on sales and market concentration across product categories in e-commerce. With the rise of online shopping, the growth of the sustainable consumer segment, and Amazon's launch of the CPF badge, our research joins the effort by documenting the causal effects of a unified third-party eco-label in the e-commerce landscape.

3. Theoretical Model

We establish a game theory model to explore the impact of the adoption of a unified green badge on demand, price, and market concentration. The model helps us understand the conditions under which badge adoption can lead to increased price, demand, and market concentration.

3.1 Sellers and Marketplace

Two competing sellers, A and B, sell two substitutable green products in the online marketplace. We follow previous research and assume these sellers' marginal production costs are zero (Zhou and Zou

2023). The two sellers determine their product prices, p_A and p_B, and pay the marketplace a percentage commission r ($0 < r < 1$) on the sale of their products. Hence, the marketplace's profits are proportional to the aggregated revenue generated by A and B. For a unit sale of product $j \in \{A, B\}$, the marketplace's profit is $r \times p_j$ and seller j's profit is $(1 - r) \times p_j$. The main analysis focuses on the case where the marketplace's commission rate r is exogenous, which is consistent with the current industry practice that commission rates seldom change even if online marketplaces have experienced many changes.[3]

The marketplace decides whether to award the unified green badge to products based on its own sustainability standards. Suppose that the sustainability level $f_j \in (0, 1]$ of product j ($j \in \{A, B\}$), which is assessed by the product's sustainability features (e.g., material) or factors in the product life cycle (e.g., carbon emission during production), is exogenously given and will not change in a short time. Without loss of generalizability, we assume that product A has a high sustainability level, with $0.5 < f_A < 1$, while product B has a low sustainability level, with $0 < f_B < 0.5$.

Additionally, the marketplace selects a threshold, denoted by $I_0 \in [0,1]$, to selectively assign a badge to a product. For the two-seller game, various badge thresholds lead to three situations. When $I_0 = 1$, neither of the two products is badged, representing a baseline case where the unified badge policy has not been implemented by the marketplace yet; when $I_0 = 0.5$, only one of the products is badged; and when, $I_0 = 0$, both products are badged. These three situations are extreme cases, as in reality, there are many products, and the badge proportion can take continuous values; it also cannot be 0 or 1. For simplicity, we consider only these three cases given the two competing products. We assume that the feature of being badged is not incorporated into the recommendation system of the marketplace. Consistently, our empirical analysis of the Amazon search data did not find evidence that a badge had an effect on the search rank, possibly because the green badge policy was a new practice (details are in Web Appendix D).

[3] The commissions for most categories have remained 15% on Amazon, 10% on eBay, and 5% on Tmall for years.

3.2 Consumers

In line with previous literature, we assume consumers are heterogeneous in their horizontal preferences toward products A and B (Singh and Vives 1984). We model this heterogeneity as two segments: the green consumer segment of size $k \in (0, 1)$ and the non-green consumer segment of size $1 - k \in (0, 1)$. Both segments are aware of both products, A and B.

Segment 1: green consumers. These consumers have a higher preference for badged products. Specifically, consumer i's preference for the two products can be represented by the following utility function (Singh and Vives 1984):

$$u_{i,G} = (\alpha_{Ai} + \rho_i G_A)q_{iA} + (\alpha_{Bi} + \rho_i G_B)q_{iB} - \frac{\beta}{2}\left(q_{iA}^2 + q_{iB}^2 + 2zq_{iA}q_{iB}\right) - p_A q_{iA} - p_B q_{iB}. \quad (1)$$

In Equation (1), q_{ij} is consumer i's consumption quantity of product j, while p_j is the price of product j set by the seller; $G_j := \mathbb{1}(f_j \geq I_0)$ is the dummy variable indicating whether product $j \in \{A, B\}$ is badged, with f_j being the sustainability level of product j, and I_0 being the badging threshold set by the marketplace, as introduced before. Coefficient α_{ji} captures consumer i's heterogeneous marginal utility from consuming product j ($j \in \{A, B\}$) based on its features, excluding sustainability-related ones, and follows a uniform distribution on $[\alpha_{j0} - c, \alpha_{j0} + c]$; ρ_i captures consumer i's heterogeneous marginal utility on the sustainability level from consuming either product, and we assume that ρ_i follows a uniform distribution on $[\rho_0 - m, \rho_0 + m]$; we further assume that $\rho_0 \gg m$ and $\alpha_{j0} \gg c$ so that consumer i's marginal utility for product features unrelated to the sustainability of product j, α_{ji}, is always greater than 0, and the sustainability level $\rho_i > 0$. We denote by α_{j0} and ρ_0 the baseline utility for product j's non-sustainability-related features and sustainability-related features, respectively. Additionally, $\beta > 0$ captures consumer i's extent of diminishing marginal utility from further consumption, and $0 < z < 1$ measures the level of substitutability or similarity between A and B for a given consumer (i.e., how a product's price will affect her purchase quantity of the other product); the larger the parameter z, the higher the similarity between two products.

To explore the effect of badge adoption on the demand, we assume that the consumers in the green segment buy only badged products and follow a two-step decision-making process before purchase. First, they maximize their utility by simultaneously deciding the purchase quantities of the two products. We look at the first-order conditions (FOCs)

$$\frac{\partial u_{i,G}}{\partial q_{iA}} = \alpha_{Ai} + \rho_i G_A - \beta(q_{iA} + zq_{iB}) - p_A = 0, \tag{2}$$

$$\frac{\partial u_{i,G}}{\partial q_{iB}} = \alpha_{Bi} + \rho_i G_B - \beta(q_{iB} + zq_{iA}) - p_B = 0. \tag{3}$$

Second, they buy only badged products and refrain from buying if the product is not badged. This assumption is consistent with the primary goal of the CPF program, which is to cater to the green segment of consumers. It follows that their purchase quantities are

$$q_{iA,G} = G_A[\alpha_{Ai} + \rho_i G_A - p_A - z(\alpha_{Bi} + \rho_i G_B - p_B)]/[\beta(1 - z^2)], \tag{4}$$

$$q_{iB,G} = G_B[\alpha_{Bi} + \rho_i G_B - p_B - z(\alpha_{Ai} + \rho_i G_A - p_A)]/[\beta(1 - z^2)]. \tag{5}$$

Segment 2: non-green consumers. The consumers in this segment do not have a preference for green or sustainability features. Thus, they differentiate products A and B based on the other features only. Specifically, consumer i's preference for the two products can be represented by the following utility function (Singh and Vives 1984):

$$u_{i,N} = \alpha_{Ai} q_{iA} + \alpha_{Bi} q_{iB} - \frac{\beta}{2}(q_{iA}^2 + q_{iB}^2 + 2zq_{iA}q_{iB}) - p_A q_{iA} - p_B q_{iB}, \tag{6}$$

where q_{ij} is consumer i's consumption quantity of product j, while p_j is the price of product j. For coefficients, α_{ji} captures consumers' heterogeneous marginal utility from consuming product j (j ∈ {A, B}) based on its non-sustainability-related features, and follows a uniform distribution on $[\alpha_{j0} - c, \alpha_{j0} + c]$; we further assume that $\alpha_{j0} \gg c$ so that the marginal utility of consumer i for the other features of product j excluding sustainability ones, α_{ij}, remains positive. Additionally, $\beta > 0$ captures the consumer's extent of diminishing marginal utility from further consumption, and $0 < z < 1$ measures the level of substitutability between A and B. Consumers buy products to maximize their utility. The FOCs

$$\frac{\partial u_{i,N}}{\partial q_{iA}} = \alpha_{Ai} - \beta(q_{iA} + zq_{iB}) - p_A = 0, \tag{7}$$

$$\frac{\partial u_{i,N}}{\partial q_{iB}} = \alpha_{Bi} - \beta(q_{iB} + zq_{iA}) - p_B = 0, \tag{8}$$

give us the optimal quantities of both products as

$$q_{iA,N} = [\alpha_{Ai} - p_A - z(\alpha_{Bi} - p_B)]/[\beta(1 - z^2)], \tag{9}$$

$$q_{iB,N} = [\alpha_{Bi} - p_B - z(\alpha_{Ai} - p_A)]/[\beta(1 - z^2)]. \tag{10}$$

3.3 Game Structure

The game has three stages (see Figure 1). In the first stage, the marketplace sets the badge threshold for green products from three choices $I_0 \in \{0, 0.5, 1\}$. In the second stage, the two sellers simultaneously set their prices, p_A and p_B. In the last stage, consumers make their purchase decisions, and profits are realized for the marketplace and the sellers.

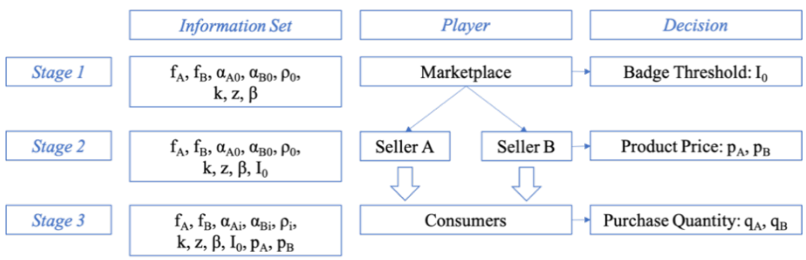


Figure 1. Three-Stage Game Structure

3.4 Equilibrium Outcome

This three-stage game has a few noteworthy equilibrium outcomes: First, under certain conditions, adopting a unified green badge leads to higher prices, increased demand, and reduced market concentration. Second, suppose the products sold by both sellers have the CPF badge, the aforementioned outcomes are ensured if the baseline utility for product features unrelated to sustainability ones significantly exceeds that for sustainability-related features. Third, when one product is badged, the sufficient condition for these outcomes is that the badge adopter's products provide higher baseline utility to consumers than the non-adopter's products.

We proceed with our analysis as follows. First, we assume that the marketplace endogenously decides the badge threshold and analyze the game with backward induction, starting from the seller's profit maximization. Given the badge threshold, the demand of product j ($j \in \{A, B\}$) is

$$D_j = k \cdot \int_{\rho_0 - m}^{\rho_0 + m} \int_{\alpha_{B0} - c}^{\alpha_{B0} + c} \int_{\alpha_{A0} - c}^{\alpha_{A0} + c} q_{ij,G} \, dF_{\alpha_{Ai}}(\alpha_{Ai}) \, dF_{\alpha_{Bi}}(\alpha_{Bi}) \, dF_{\rho_i}(\rho_i)$$

$$+(1-k)\cdot\int_{\alpha_0-c}^{\alpha_0+c}\int_{\alpha_0-c}^{\alpha_0+c} q_{ij,N}\, dF_{\alpha_{Ai}}(\alpha_{Ai})\, dF_{\alpha_{Bi}}(\alpha_{Bi}), \tag{11}$$

where the two terms are the expected purchase quantity from segment 1 (green consumers) and segment 2 (non-green consumers), respectively.

Next, we derive the profit-maximizing price and demand of the two sellers for the three choices of the badging threshold set by the marketplace in the first stage.

Situation 1: badge threshold = 1. This baseline situation where no product is badged is the same as the situation where the marketplace has not implemented the unified badge policy. Therefore, $G_j = 0$, and demand comes only from the non-green segment. The demand for product A is

$$D_A^{(1)} = \frac{1-k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} - p_A + zp_B], \tag{12}$$

and the demand for product B is

$$D_B^{(1)} = \frac{1-k}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - p_B + zp_A]. \tag{13}$$

Thus, the profit derived from selling product A is

$$\Pi_A^{(1)} = (1-r)p_A \cdot \frac{1-k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} - p_A + zp_B], \tag{14}$$

and the profit derived from selling product B is

$$\Pi_B^{(1)} = (1-r)p_B \cdot \frac{1-k}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - p_B + zp_A]. \tag{15}$$

When maximizing the profit functions of products A and B simultaneously, there is only one root when FOC = 0. Therefore, we have

$$p_A^{*,(1)} = \frac{\alpha_{A0} - z\alpha_{B0} + zp_B^{*,(1)}}{2},\ p_B^{*,(1)} = \frac{\alpha_{B0} - z\alpha_{A0} + zp_A^{*,(1)}}{2}, \tag{16}$$

which thus gives us the final expression for optimal prices and demand as summarized in Lemma 1.

Lemma 1. *When the badging threshold is 1, no product is badged, the profit-maximizing subgame equilibrium has optimal prices as* $p_A^{*,(1)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{A0} - z\alpha_{B0}]$, $p_B^{*,(1)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{B0} - z\alpha_{A0}]$, *and the demand for products A and B of this subgame perfect equilibrium is* $D_A^{*,(1)} =$

$\frac{1-k}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{A0} - z\alpha_{B0}]$ *and* $D_B^{*,(1)} = \frac{1-k}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{B0} - z\alpha_{A0}]$, *respectively.*

Note that to make sure that prices are nonnegative, we need to add the constraint

$$\alpha_{A0} > \frac{z}{2-z^2}\alpha_{B0} > \left(\frac{z}{2-z^2}\right)^2 \alpha_{A0}, \tag{17}$$

which gives us the condition that $z < 1$. Given that we assume $0 < z < 1$, as it measures the level of substitutability between products A and B for a given consumer, such a condition is already satisfied.

Situation 2: badge threshold = 0.5. This is an asymmetric structure where only product A is badged, as $f_A > I_0$ while $f_B < I_0$. This means that $G_A = 1$ while $G_B = 0$, and the utility functions of both consumer segments are different due to different consumer preferences for badged products. The demand for product A is

$$D_A^{(2)} = \frac{k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + \rho_0 - p_A + zp_B] + \frac{1-k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} - p_A + zp_B]$$

$$= \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + k\rho_0 - p_A + zp_B], \tag{18}$$

whereas the demand for product B is

$$D_B^{(2)} = \frac{1-k}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - p_B + zp_A]. \tag{19}$$

Thus, the profit derived from selling product A is

$$\Pi_A^{(2)} = (1-r)p_A \cdot \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + k\rho_0 - p_A + zp_B], \tag{20}$$

and, the profit derived from selling product B is

$$\Pi_B^{(2)} = (1-r)p_B \cdot \frac{1-k}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} - p_B + zp_A]. \tag{21}$$

We maximize the profit derived from selling products A and B simultaneously to obtain their optimal prices,

$$p_A^{*,(2)} = \frac{\alpha_{A0} - z\alpha_{B0} + k\rho_0 + zp_B^{*,(2)}}{2}, p_B^{*,(2)} = \frac{\alpha_{B0} - z\alpha_{A0} + zp_A^{*,(2)}}{2}, \tag{22}$$

which thus gives us the final expression for optimal prices and demand as summarized in Lemma 2.

Lemma 2. *When the badging threshold is 0.5, only product A is badged, and the profit-maximizing subgame equilibrium has optimal price as* $p_A^{*,(2)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{A0} - z\alpha_{B0} + 2k\rho_0]$, $p_B^{*,(2)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{B0} - z\alpha_{A0} + zk\rho_0]$. *The demand for products A and B of this subgame perfect equilibrium is* $D_A^{*,(2)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{A0} - z\alpha_{B0} - (1-z^2)k\rho_0]$ *and* $D_B^{*,(2)} = \frac{1-k}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{B0} - z\alpha_{A0} + zk\rho_0], respectively.$

Based on Lemma 1 and Lemma 2, we can see that the optimal prices in situation 2 are larger than those in situation 1 and thus larger than 0. By comparing the price difference between situation 1 and situation 2,

$$p_A^{*,(2)} - p_A^{*,(1)} = \frac{2k\rho_0}{4-z^2}, p_B^{*,(2)} - p_B^{*,(1)} = \frac{zk\rho_0}{4-z^2}, \tag{23}$$

we can conclude that the price increase in situation 2 is due to the consumers' preference for sustainability features, which can be understood as the price premium caused by green features.

If the demand for product A in situation 2 is larger than that in situation 1,

$$D_A^{*,(2)} - D_A^{*,(1)} = \frac{k}{\beta(1-z^2)(4-z^2)}[(2-z^2)\alpha_{A0} - z\alpha_{B0} - (1-z^2)\rho_0] > 0, \tag{24}$$

where the first two terms represent the influx of green consumers, and the second term represents the loss of non-green consumers due to higher prices. This gives us the condition that

$$(2-z^2)\alpha_{A0} > z\alpha_{B0} + (1-z^2)\rho_0. \tag{25}$$

If we want the demand for product B in situation 2 to be larger than that in situation 1, we have

$$D_B^{*,(2)} - D_B^{*,(1)} = \frac{(1-k)zk\rho_0}{\beta(1-z^2)(4-z^2)} > 0, \tag{26}$$

which holds true under all conditions. This phenomenon can be attributed to the fact that product B holds a greater price advantage in situation 2 compared to situation 1, as product B's price does not increase as significantly as product A's in the former scenario. This gives us proposition 1.

Proposition 1. *The prices and demand in situation 2 (one product badged) are larger than those in situation 1 (no product badged) under the sufficient condition that the baseline preference for non-sustainability-related features of the badged product is sufficiently high compared with that of the* *unbadged product; also, the baseline utility for non- sustainability-related features of the badged product is sufficiently large.*

An intuitive explanation for this proposition is that badge adoption has a trade-off effect on the demand of green consumers. On the one hand, the demand might increase because the utility of green consumers increases if the product has sustainable features; on the other hand, the demand might decrease because there is a price premium for badged products that drives away the non-green segment of consumers. Therefore, the baseline utility for non-sustainability-related features of product A needs to be large enough so that the gain in green consumers outweighs any loss of non-green consumers due to higher prices.

On the other hand, if the demand for product A in situation 2 is smaller than that in situation 1, then

$$(2 - z^2)\alpha_{A0} < z\alpha_{B0} + (1 - z^2)\rho_0, \tag{27}$$

which can be achieved under the sufficient condition that the price premium caused by the high preference for the green badge is high, such that the term $(1 - z^2)\rho_0$ is high, or a fierce competition from product B such that the term $z\alpha_{B0}$ is high.

Situation 3: badge threshold = 0. In this situation both products are badged (i.e., $G_A = 1$ and $G_B = 1$), meaning that the demand for products A and B from both consumer segments is nonzero. The demand for product A is

$$D_A^{(3)} = \frac{k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (1 - z)\rho_0 - p_A + zp_B] + \frac{1-k}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} - p_A + zp_B]$$

$$= \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (1 - z)k\rho_0 - p_A + zp_B], \tag{28}$$

and the demand for product B is

$$D_B^{(3)} = \frac{1}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} + (1 - z)k\rho_0 - p_B + zp_A]. \tag{29}$$

Thus, the profit derived from selling product A is

$$\Pi_A^{(3)} = (1 - r)p_A \cdot \frac{1}{\beta(1-z^2)}[\alpha_{A0} - z\alpha_{B0} + (1 - z)k\rho_0 - p_A + zp_B], \tag{30}$$

and the profit derived from selling product B is

$$\Pi_B^{(3)} = (1-r)p_B \cdot \frac{1}{\beta(1-z^2)}[\alpha_{B0} - z\alpha_{A0} + (1-z)k\rho_0 - p_B + zp_A]. \quad (31)$$

Profit maximization implies that the optimal prices of products A and B are

$$p_A^{*,(3)} = \frac{\alpha_{A0} - z\alpha_{B0} + (1-z)k\rho_0 + zp_B^{*,(3)}}{2},\ p_B^{*,(3)} = \frac{\alpha_{B0} - z\alpha_{A0} + (1-z)k\rho_0 + zp_A^{*,(3)}}{2}, \quad (32)$$

which thus gives us the final expression for optimal prices and demand as summarized in Lemma 3.

Lemma 3. *When the badging threshold is 0, both products are badged, and the profit-maximizing subgame equilibrium has optimal price* $p_A^{*,(3)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{A0} - z\alpha_{B0} + (2-2z)k\rho_0]$, $p_B^{*,(3)} = \frac{1}{4-z^2}[(2-z^2)\alpha_{B0} - z\alpha_{A0} + (2-2z)k\rho_0]$. *The demand for products A and B of this subgame perfect equilibrium is* $D_A^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(z+4-2z^2)\alpha_{A0} - (2z+2-z^2)\alpha_{B0} + (1-z)(2z+2-z^2)k\rho_0]$ and $D_B^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(z+4-2z^2)\alpha_{B0} - (2z+2-z^2)\alpha_{A0} + (1-z)(2z+2-z^2)k\rho_0], respectively.$

Given that 0 < z < 1, as it measures the level of substitutability between products A and B for a given consumer, such a condition makes sure that the prices are positive. Based on Lemmas 1 and 3, the sub-equilibrium prices in situation 3 are obviously larger than those in situation 1, owing to the influx of green consumers. Next, if we want the demand for product A in situation 3 to be at least larger than that in situation 1, we have the following conditions for product A,

$$(1-k)[(2-z^2)\alpha_{A0} - z\alpha_{B0}] < [(2-z^2)\alpha_{A0} - z\alpha_{B0} - (z^2-2z-2)(1-z)k\rho_0],$$

which gives us the condition that

$$z\alpha_{B0} + (2+2z-z^2)(1-z)\rho_0 < (2-z^2)\alpha_{A0}. \quad (33)$$

Similarly, we have the following condition for product B,

$$z\alpha_{A0} + (2+2z-z^2)(1-z)\rho_0 < (2-z^2)\alpha_{B0}, \quad (34)$$

based on which we discuss the sufficient condition for increased price and demand.

Proposition 2. *One sufficient condition that prices and demand in situation 3 (both products badged) are larger than those in situation 1 (no product badged) is that the baseline utility for non-*

sustainability-related features of products A and B is sufficiently larger than the baseline utility for sustainability-related features.

As a numerical example, suppose that $\alpha_{A0} = \alpha_{B0} = 4\rho_0 > 1$; then the conditions are simplified to $(2 + 2z - z^2) < 4(2 + z)$, which is always true, as $(z + 1)^2 + 5 > 0$. The intuition follows the same logic when we compare situation 2 with situation 1: Demand might increase due to the higher utility green consumers derive from sustainable features. However, demand could also decrease because the price premium for badged products may deter non-green consumers from buying. Therefore, the baseline utility of badged products' non-sustainability features must be substantial enough for the gain in green consumers to offset any loss of non-green consumers caused by the higher prices.

The sufficient condition for the demand for product A in situation 3 to be smaller than that in situation 1 is

$$z\alpha_{B0} + (2 + 2z - z^2)(1 - z)\rho_0 > (2 - z^2)\alpha_{A0}, \tag{35}$$

which can be achieved under the sufficient condition that a demand decrease is caused by the high price premium for product A's sustainability features or a fierce competition from product B. The same applies to the demand for product B.

Finally, if the price and demand for products A and B in situation 2 are larger than those in situation 3, such that the medium-level badging threshold is the optimal choice, then the following conditions hold. For price, given that $p_A^{*,(2)} - p_A^{*,(3)} = 2zk\rho_0 > 0$ and $p_B^{*,(2)} - p_B^{*,(3)} = (3z - 2)k\rho$, the sufficient condition for both prices of products A and B being higher in situation 2 than in situation 3 is that $0.67 < z < 1$.

For demand, we have that

$$D_A^{*,(2)} - D_A^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(z + 1)(z - 2)(\alpha_{A0} - \alpha_{B0}) - (3 - 4z^2 + z^3)k\rho_0], \tag{36}$$

$$D_B^{*,(2)} - D_B^{*,(3)} = \frac{1}{\beta(1-z^2)(4-z^2)}[(z + 1)(z - 2)(\alpha_{B0} - \alpha_{A0}) - (2 - 3z^2 + z^3 - z)k\rho_0], \tag{37}$$

which gives the range of $\alpha_{A0} - \alpha_{B0}$ if we want $D_A^{*,(2)} - D_A^{*,(3)} > 0$ and $D_B^{*,(2)} - D_B^{*,(3)} > 0$,

$$\frac{3-4z^2+z^3}{(z+1)(z-2)} < \alpha_{A0} - \alpha_{B0} < \frac{3z^2-z^3+z-2}{(z+1)(z-2)}. \quad (38)$$

Thus, as long as there exists *z* that satisfies such a condition, there exists a situation where demand in situation 2 is higher than in situation 3. By solving the inequality

$$\frac{3-4z^2+z^3}{(z+1)(z-2)} > \frac{3z^2-z^3+z-2}{(z+1)(z-2)},$$

we have that $g(z) = -2z^3 + 7z^2 + z - 5 > 0$ when $0 < z < 0.89$, which if intersect within the range of $0.67 < z < 1$, we have that $0.67 < z < 0.89$. When $0.67 < z < 0.89$, we have that $\frac{3z^2-z^3+z-2}{(z+1)(z-2)} < \frac{3-4z^2+z^3}{(z+1)(z-2)} < 0$. Therefore, a sufficient condition for prices and demands of each product in situation 2 to be higher than those in situation 3 is that $0.67 < z < 0.89$ and that

$$\frac{3-4z^2+z^3}{(z+1)(z-2)} < \alpha_{A0} - \alpha_{B0} < \frac{3z^2-z^3+z-2}{(z+1)(z-2)} < 0.$$

Combining this condition with the sufficient condition when the price and demand in situation 2 are higher than those in situation 1, we have the following proposition.

Proposition 3*. Profit maximization for the marketplace and both sellers can be achieved*[4] *at the medium level of the badging threshold when the following three sufficient conditions hold: (1)* $(2 - z^2)\alpha_{A0} > z\alpha_{B0} + (1 - z^2)\rho_0$*, (2)* $\frac{3-4z^2+z^3}{(z+1)(z-2)} < \alpha_{A0} - \alpha_{B0} < \frac{3z^2-z^3+z-2}{(z+1)(z-2)} < 0$*, (3) 0.67 < z < 0.89; that is, the baseline utility for non-sustainability-related features of the unbadged product is slightly higher than the baseline utility for non-sustainability-related features of the badged product A; the baseline utility for sustainability features is reasonably smaller than the baseline utility for non-sustainability-related features of the badged product; and the product differentiation between two sellers is sufficiently high.*

For instance, if z = 0.8, the condition becomes $1.36\alpha_{A0} > 0.8\alpha_{B0} + 0.36\rho_0$, which can be simplified as $\alpha_{A0} > 0.59\alpha_{B0} + 0.26\rho_0$ and $\alpha_{B0} - 0.44 < \alpha_{A0} < \alpha_{B0} - 0.10$. Under this condition, as long as

[4] Since we are looking for a sufficient condition of increased profit for both the platform and the sellers, we compare price and demand separately so that the highest profit is achieved when both price and demand are highest among the three situations.

$\alpha_{B0} > 0.23 + 0.64\rho_0$, a valid α_{A0} exists. If, for instance, $\alpha_{B0} = 2$ and $\rho_0 = 0.5$, then α_{A0} can take any value in the range of (1.56, 1.90).

Comparisons on market concentration. Finally, we look at market concentration under three situations. To capture market concentration, we calculate the Herfindahl-Hirschman index (HHI), which is the sum of the squares of the market shares of all firms operating in a particular market; the indicator has been widely adopted in past research (Kelly 1981). The HHI gives higher weights to larger firms, with higher values indicating greater market concentration, and is calculated as follows:

$$\text{HHI} = \sum_{i=1}^{n} s_i^2, \tag{39}$$

where n is the number of firms in the market, and s_i is the market share of firm i. The market share is

$$MS_A = \frac{D_A^*}{D_A^* + D_B^*},\ MS_B = \frac{D_B^*}{D_A^* + D_B^*},\ MS_A + MS_B = 1, \tag{40}$$

based on which we can calculate the HHI as

$$HHI^* = (MS_A)^2 + (MS_B)^2 = (MS_A)^2 + (1 - MS_A)^2 \geq 0.25, \tag{41}$$

which takes the minimum when $MS_A = MS_B$, or, in other words, when $D_A^* = D_B^*$. Based on the U-shape curve between HHI and MS_A (Figure 2), we can now derive that the closer MS_A is to 0.5, the smaller the HHI, and thus the lower the market concentration.

By looking at the subgame perfect equilibrium under three situations, we have

$$|MS_A^{(1)} - \tfrac{1}{2}| = \frac{|D_A^{*,(1)} - D_B^{*,(1)}|}{2(D_A^{*,(1)} + D_B^{*,(1)})} = \frac{(1+z)(2-z)|\alpha_{A0} - \alpha_{B0}|}{2(1-z)(2+z)(\alpha_{A0} + \alpha_{B0})}, \tag{42}$$

$$|MS_A^{(2)} - \tfrac{1}{2}| = \frac{|D_A^{*,(2)} - D_B^{*,(2)}|}{2(D_A^{*,(2)} + D_B^{*,(2)})} = \frac{|(2-z^2+(1-k)z)\alpha_{A0} - (z+(1-k)(2-z^2))\alpha_{B0} - (1-z^2+(1-k)z)k\rho_0|}{2[(2-z^2-(1-k)z)\alpha_{A0} + ((1-k)(2-z^2)-z)\alpha_{B0} + ((1-k)z+z^2-1)k\rho_0]}, \tag{43}$$

$$|MS_A^{(3)} - \tfrac{1}{2}| = \frac{|D_A^{*,(3)} - D_B^{*,(3)}|}{2(D_A^{*,(3)} + D_B^{*,(3)})} = \frac{(1+z)(2-z)|\alpha_{A0} - \alpha_{B0}|}{2(1-z)(2+z)(\alpha_{A0} + \alpha_{B0}) + 4(2+2z-z^2)(1-z)k\rho_0}. \tag{44}$$

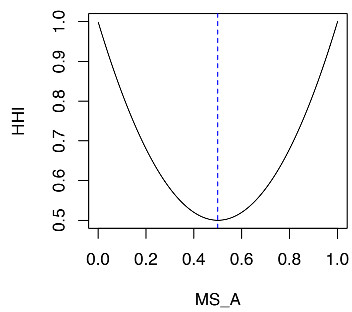


Figure 2. U-shape Relationship between Market Share of Product A (MS_A) and Market Concentration (Herfindahl-Hirschman Index, HHI)

First, we compare situation 2 with situation 1. If the condition that the demand for product A is larger in situation 2 than that in situation 1 is satisfied,

$$(2 - z^2)\alpha_{A0} > z\alpha_{B0} + (1 - z^2)\rho_0, \tag{45}$$

we can see that the $|MS_A^{(2)} - 1/2|$ can be smaller than $|MS_A^{(1)} - 1/2|$ if α_{A0} is much larger than α_{B0} to compensate for the price premium of being badged in the above condition. This gives proposition 4 where we discuss sufficient conditions for increased price, demand, and market competition.

Proposition 4. *Increased price, demand, and market competition in situation 2 (one product badged) as compared with situation 1 (no product badged) can be achieved if the baseline utility for non-sustainability-related features of the badged product is much larger than that of the unbadged product.*

As a numerical example, if we set $\alpha_{A0} = 64$, $\alpha_{B0} = 8$, and $\rho_0 = 16$, $z = 0.5$, then when k takes a value from 0 to 1, we always have that $|MS_A^{(2)} - 1/2|$ is smaller than $|MS_A^{(1)} - 1/2|$, with $|MS_A^{(2)} - 1/2|$ decreasing as the green segment becomes larger (Figure 3).

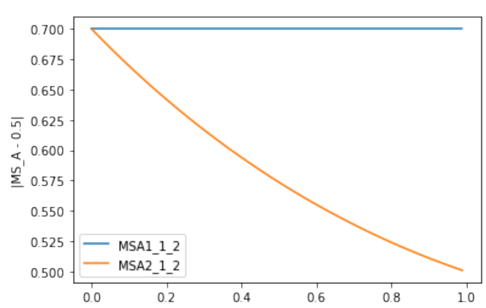


Figure 3. Numerical Simulation Comparing the distance of market share of product A in situation 1 and 2 with Different Sizes of the Green Segment

Second, we compare situation 3 with situation 1. The only difference between $|MS_A^{(3)} - 1/2|$ and $|MS_A^{(1)} - 1/2|$ lies in the denominator. The denominator of $|MS_A^{(3)} - 1/2|$ is always larger than that of $|MS_A^{(1)} - 1/2|$ as the coefficient of $k\rho_0$ is always larger than 0. This gives proposition 5 where we discuss sufficient conditions for increased price, demand, and market competition, based on Proposition 2.

Proposition 5*. Comparing situation 3 (both products badged) with situation 1 (no product badged), to achieve increased price, demand, and decreased market concentration in situation 3, one sufficient condition is that the value of the baseline utility for non-sustainability-related features of both products is sufficiently larger than the value of the baseline utility for sustainability-related features.*

To summarize, if the equilibrium represents increased price, demand, and decreased market concentration when the unified green badge policy is implemented as compared to when no product is badged, then the baseline utility of badged products' non-sustainability features must be substantial enough for the gain in green consumers to offset any loss of non-green consumers due to higher prices.

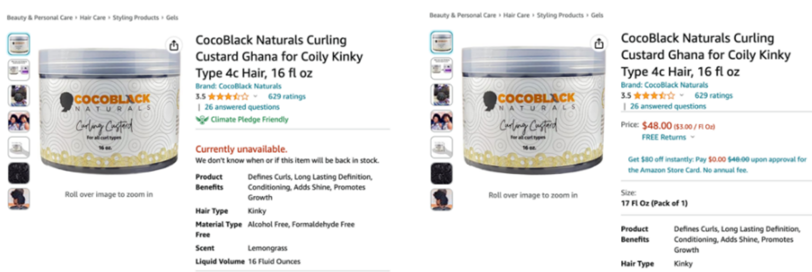
4. Data

To explore the causal impact of a unified green badge adoption on product demand, price, and market concentration, and to assess the potential for a win-win-win outcome as predicted by our theoretical model, we collected 6.5 months of data from Amazon on 6,606 distinct products across 8 categories and 20 subcategories. We found 35.4% of the products in our data were badged with the CPF label at least once.

4.1 Price and Demand

We collect daily data on 14,000 products across eight categories with a high percentage of green products on Amazon from March 1, 2023, to September 15, 2023. These categories are Beauty & Personal Care, Health & Household, Grocery & Gourmet Food, Clothing, Shoes & Jewelry, Sports & Outdoors, Office Products, Electronics, and Toys & Games. After removing incomplete observations and products with fewer than 135 days of data (out of 193 days), we restricted our sample to 6,606 distinct products. Sales rank, a widely used metric for quantifying demand on Amazon, was our primary measure, with lower ranks indicating higher sales. We identified products using their Amazon Standard Identification Number (ASIN), a 10-character alphanumeric identifier.[5] Table W1 in Web Appendix A details the range of products by type and subtype and the presence of green certifications.

For each product, we collected the following metrics daily: price, rating count, mean positive scores of reviews, and mean rating count. In addition, we collected the following metrics weekly: mean compound valence score, length, richness, and readability of product description; and colorfulness, brightness, symmetry, average face count, technical quality score, and aesthetics quality score of product images. During the period of our data collection, we observed instances of treatment reversal, where products experienced changes in their badge status. Specifically, some products initially did not have a badge but were later awarded one, while others initially had a badge but subsequently had it removed. This dynamic nature of badge allocation provided a unique opportunity to observe the effects of these changes. For example, in Figure 4 (a) and (b), a product initially possessed a badge but had it removed on May 18, 2023. Web Appendix B presents the history of CPF badge status for a random sample.



(a) Product Screenshot on May 17, 2023 (b) Product Screenshot on May 20, 2023

Figure 4. Examples of Product Before Badge Removal (Left) and After Badge Removal (Right)

4.2 Product Images

We collected weekly product images from March 1, 2023, to September 15, 2023, on the 6,606 distinct products in our demand dataset. We used computer vision techniques and deep learning models to extract interpretable product features from unstructured image data. The list of features we extracted

[5] See detailed examples and introduction at https://developer.amazon.com/docs/mobile-associates/mas-finding-product-id.html#:~:text=You%20can%20find%20the%20ASIN,URL%20of%20the%20product%20page.

comprises those that have been proven critical for consumer decisions: (1) colorfulness (2) brightness, (3) visual balance, (4) visual complexity, (5) image quality, and (6) face count.

Colorfulness measures how chromatic (any color lacking white, gray, or black) the perceived color of an area appears to be (Fairchild 2013).

The general *brightness* of an image holds importance. Ample illumination is crucial in rendering image content clear to viewers, as the information within images is communicated via pixel brightness (Gorn et al. 1997).

Visual balance can be understood as the symmetry of an image's visual elements (in this case, intensity and color). Visually balanced real estate images give viewers a feeling of order and tidiness, minimizing the cognitive demand required to process the images (Kreitler and Kreitler 1972).

Visual complexity can impact both liking and usability in advertisements. We use edge density as a proxy for visual complexity ((Pieters, Wedel, and Batra 2010).

Image quality can have a significant impact on consumer decisions and product demand, especially in online transactions (Zhang et al., 2022). We used Neural IMage Assessment (NIMA) to assess the image quality of product images (Talebi and Milanfar 2018; Ceylan, Diehl, and Proserpio, 2023).

Face count can be understood as the approximate number of human models employed by the brand in the advertisement. Past literature has shown the significant impact that a human model, especially an attractive one, in a product advertisement has on sales (Baker and Churchill 1977).

4.3 Product Descriptions and Customer Reviews

We also collected weekly product descriptions from March 1, 2023, to September 15, 2023, on the 6,606 distinct products in the demand dataset. We used advanced natural language processing techniques and deep learning models to extract interpretable product features from unstructured text data. The list of features we extracted comprises (1) valence, (2) richness, and (3) readability. Note that we tracked only the reviews' characteristics, not the reviewers'.

Valence refers to the emotion conveyed in the product description, indicating whether the language evokes positive, negative, or neutral emotions among consumers.

Richness refers to lexical richness, and it encompasses the breadth and diversity of vocabulary within a given text. This aspect of language can be a significant indicator of various factors, including writing quality and vocabulary knowledge (McCarthy and Jarvis 2007).

Past research has shown that more readable descriptions are more memorable (Reczek et al. 2018). We used the Flesch Reading Ease Score test to assess a text's *readability* by analyzing word, syllable, and sentence counts.

Additionally, we collected all customer reviews listed on the "See more reviews" page as of September 16, 2023. Since Amazon frequently detects and deletes fake reviews (He, Hollenbeck, and Proserpio 2022), we did not rely solely on the rating count and score displayed on the product detail page. Instead, we calculated the newly added and total review counts and computed the average rating score based on authentic (non-deleted) reviews. We also performed sentiment analysis on customer reviews using Vader's sentiment analyzer (Hutto and Gilbert 2014), as previous research has shown that review sentiment impacts demand (Liu, Lee, and Srinivasan 2019). In addition, we counted word occurrences related to sustainability[6] as a proxy for the extent to which the review focuses on sustainability; we also counted word occurrences pertaining to packaging and return[7] as a proxy for the extent to which the review is written for the delivery process.

4.4 Multimodal Vector Representations

To eliminate concerns about the inadequate selection of image and description features as introduced in subsections Product Images and Product Descriptions and Customer Reviews, we alternatively extracted non-interpretable vector representations from a state-of-the-art multimodal machine learning model, Contrastive Language–Image Pre-training (CLIP; Radford et al. 2021). We chose CLIP mainly because its model demonstrates enhanced flexibility and generality compared to conventional ImageNet models. CLIP scales a straightforward pre-training task to attain competitive zero-shot performance

[6] The set of strings we used for calculating occurrences on sustainability-related topics include *eco*, *climate*, *environment*, *green*, *renew*, *preserv*, *endur*, *earth*, *bio*, and *sustain*.
[7] The set of strings we used for calculating occurrences on packaging-and-return related topics include *packag*, *retrun*, *parcel*, *box*, *empty*, *ship*, and *deliver*.

across diverse image classification datasets.

Figure 5 shows the three steps involved in extracting the image and description vector representations. First, we used the NLTK summarizer to shorten the long descriptions given the CLIP model has a length restriction on the text of 77 characters. Second, we inputted the image into the image feature encoder and the summarized description into the text feature encoder; both encoders have the backbone structure of ViT-B/32 Transformer architecture (Radford et al. 2021). Third, after obtaining a 512-dimensional vector representation from the image feature encoder and a 512-dimensional vector representation from the text feature encoder for all images and descriptions, we used principal component analysis to conduct dimension reduction respectively for images and descriptions. We obtained a 10-dimensional vector representation for each product image and a 10-dimensional vector representation for each (summarized) product description.

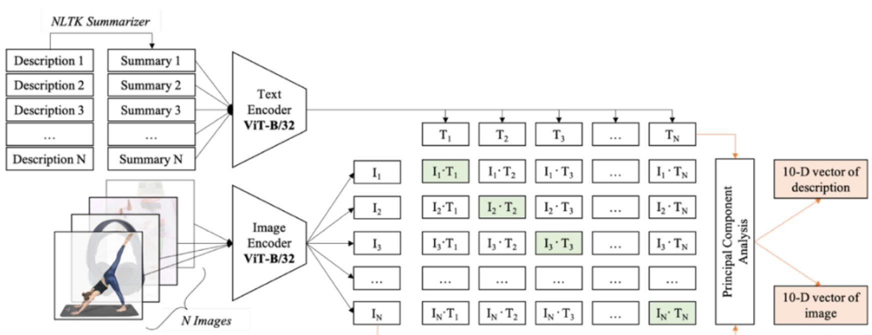


Figure 5. CLIP Framework for Extracting Image and Textual Vector Representations

5. Empirical Methods

We employ the interactive fixed effects counterfactual (IFEct) estimator (Liu, Wang, and Xu 2022) for inferring the causal impact of a CPF badge on a set of critical business outcomes: product demand, price, and market concentration. The IFEct approach fulfills our research objective for two reasons. One, it allows for time-varying covariates, which is needed because there are many time-varying features to be controlled, including price, rating score, product image, and description features. Two, it allows for treatment reversal, which appears in our data; some products did not have a badge initially but were later awarded one, while others that initially had a badge subsequently had it removed.

5.1 The IFEct Framework for Identifying the Causal Effects of the CPF Badge

Consider a balanced sample that consists of N units observed over T periods. Here D_{it} denotes the dummy variable representing being treated and 0 otherwise; Y_{it} is the outcome variable, with $Y_{it}(0)$ denoting the outcome when it is not treated and $Y_{it}(1)$ when it is treated; $\boldsymbol{X}_{it}$ is a vector of exogenous covariates; $\boldsymbol{U}_{it}$ is a vector of unobservable covariates; and ε_{it} is the idiosyncratic error term.

We are interested in quantifying the average treatment effect on the treated units (treatment status has changed at least once during the observed time window):

$$ATT = \mathbb{E}[\delta_{it}|D_{it} = 1, C_i = 1], \tag{46}$$

where $\delta_{it} = Y_{it}(1) - Y_{it}(0)$; $C_i = 1$ if unit i has the treatment status changed at least once, i.e., $\exists t, s \in \{1,2, \dots, T\}$ such that $D_{it} = 1, D_{is} = 0$; otherwise, $C_i = 0$.

We follow Liu, Wang, and Xu (2022) and make three identification assumptions: no temporal or spatial interaction, strict exogeneity, and existence of a low dimensional decomposition. Detailed discussions are in Web Appendix C.

5.2 Augmented Factors and Relaxation on Exogeneity

When unobserved time-varying confounders exist, there is concern about the endogeneity issue, and thus the exogeneity assumption will not hold. A couple of authors have proposed using factor-augmented models to relax the strict exogeneity assumption (Bai 2009; Xu 2017). Among them, IFEct models the response surface of untreated potential outcomes using a factor-augmented model. We assume that

$$Y_{it}(0) = \boldsymbol{X'}_{it}\beta + \alpha_i + \tau_t + \lambda'_i f_t + \varepsilon_{it}, \tag{47}$$

where $f_t = [f_{1t}, \dots, f_{rt}]' \in \mathbb{R}^{r\times 1}$ is a vector of unobserved common factors, and $\lambda_i = [\lambda_{i1}, \dots, \lambda_{ir}]' \in \mathbb{R}^{r\times 1}$ is a vector of unknown factor loadings. We assume that this factor component takes a linear, additive form $\lambda'_i f_t = \sum_{k=1}^{r} \lambda_{ik} f_{kt}$. In general, as long as an unobserved random variable can be decomposed into a multiplicative form, it can be absorbed by $\lambda'_i f_t$. Here $\boldsymbol{X'}_{it}$ is a vector of covariates; α_i is unit fixed effect, while τ_t is time fixed effect; and ε_{it} is a matrix of idiosyncratic errors. See an illustration in Figure 6. More detailed steps in the estimation algorithm for the IFEct estimator are in Web Appendix C.

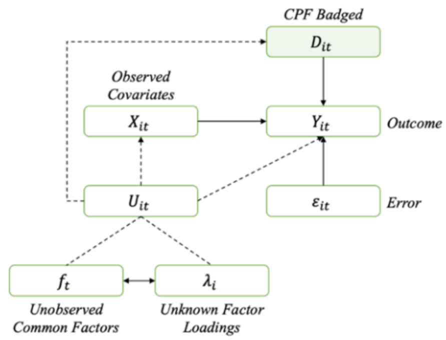


Figure 6. A Directed Graphical Model Illustration of the IFEct Estimator

Lastly, we have thoroughly assessed potential biases and confounders, including a pre-trend analysis, placebo tests, and evaluations for any carryover effects, bolstering the credibility of our findings. Additionally, we investigated the CPF badge's influence on search rankings on the platform and did not find evidence that a CPF badge alters product search rankings. Details are in Web Appendix D.

6. Empirical Results

We present our findings on the causal effects of adopting the CPF badge on product demand, pricing strategies, and market concentration. This analysis tests our three main hypotheses.

6.1 CPF Badge Adoption Leads to Increased Demand

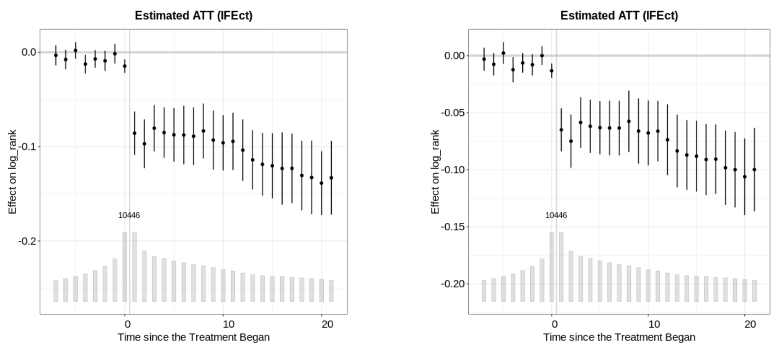
We first explore the causal impact of CPF on demand, measured as log(sales rank) (He, Hollenbeck, and Proserpio 2022; Park, Xie, and Xie 2023). We run two models with different sets of covariates. Model (1) includes only price and rating count. Model (2) adds more variables: price, rating count, mean positive review scores, mean rating count, mean compound valence score, and various features of product descriptions (length, richness, readability) and images (colorfulness, brightness, symmetry, average face count, technical quality, and aesthetic quality). We used four augmented factors to minimize root mean square error (RMSE). The ATT and estimated coefficients for each covariate are shown in Table 1, with time variation in the average treatment effect for both models illustrated in Figure 7.

We observe a negative and significant impact of adopting the CPF badge on sales rank, indicating a positive effect of this badge on sales volume. This result remains consistent even after controlling for various product features in Model (2). However, the effect size is smaller in Model (2) compared to Model (1), suggesting that product images and descriptions may significantly influence how the CPF badge affects demand. Diagnostics for Model (2) in Table 1 with the sales rank as DV, including tests for pre-trend, placebo effect, and carryover effect, are detailed in Web Appendix D.

Table 1. Estimation Results of Causal Impact of CPF Badge on Demand

DV: Log (sales rank)	**Model (1)**		**Model (2)**	
	ATT	p	ATT	p
Treated observations equally weighted	-0.1394	0.0000	-0.1041	0.0000
Treated units equally weighted	-0.1317	0.0000	-0.0974	0.0000
	$\hat{\beta}$	p	$\hat{\beta}$	p
Log (price + 1)	0.1060	0.0000	0.1279	0.0000
Log (rate # + 1)	-0.4897	0.0000	-0.5219	0.0000
Mean review positivity			0.4311	0.0872
Mean rating score			0.0479	0.0769
Log (sustainability topic + 1)			0.2781	0.0000
Log (packaging topic + 1)			0.0058	0.8441
Description valence			-0.1099	0.0877
Log (description length + 1)			-0.1174	0.0013
Description richness			-0.3331	0.0777
Description readability			-0.0020	0.0112
Colorfulness			0.0002	0.7816
Brightness			0.1305	0.4464
Symmetry			0.0000	0.4629
Face #			-0.1851	0.0679
Image technical score			0.0851	0.1295
Image aesthetic score			0.0304	0.7494
RMSE	0.3440		0.3461	
Obs. #	1274958		1274958	

Notes. For identification purposes, units whose number of untreated periods is less than 5 are dropped automatically.



(a) ATT in Model (1) (b) ATT in Model (2)

Figure 7. ATT of Having the CPF Badge on Demand from Model (1) [a] and Model (2) [b]

Notes. The gray bar denotes number of units at the *t* period after treatment. DV is sales rank; higher rank indicates lower demand.

We analyzed the average treatment effect across product categories using the same specification as in Model (2), controlling for various product features, with results presented in Table 2. The analysis shows that adopting the CPF badge had a positive and significant effect on product demand in six categories. However, in three categories—Clothing, Shoes & Jewelry, Health & Household, and Sports & Outdoors—the CPF badge did not significantly impact demand. This lack of impact could be due to a trade-off between the positive and negative aspects of eco-labeling in these categories.

Table 2. Estimation Results of Causal Impact of CPF Badge on Demand by Category

Category	ATT	p	ATT	p	RMSE	Obs. #
	Treated observations equally weighted		*Treated units equally weighted*			
Beauty & Personal Care	-0.0771	0.0009	-0.0724	0.0019	0.3252	235846
Clothing, Shoes & Jewelry	-0.1405	0.1161	-0.1166	0.2339	0.2345	109238
Electronics	-0.1561	0.0000	-0.1592	0.0000	0.3238	203229
Grocery & Gourmet Food	-0.0454	0.1745	-0.0550	0.0861	0.3490	161541
Health & Household	0.0653	0.1300	0.0590	0.1214	0.4338	251479
Office Products	-0.0936	0.0000	-0.0677	0.0001	0.2771	151505
Sports & Outdoors	0.0654	0.3560	0.0776	0.8169	0.3271	27599
Toys & Games	-0.0903	0.0182	-0.0874	0.0160	0.2693	134521

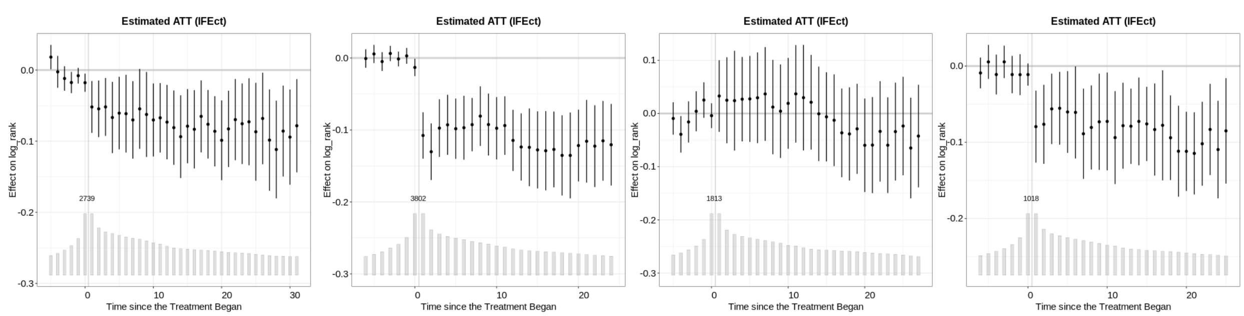
Notes. For identification purposes, units whose number of untreated periods is less than 5 are dropped automatically.

Additionally, we also looked at the effect of the CPF badge on demand for different demographic groups, including age (young vs. old) and gender group (female vs. male), where the products were selected to represent the purchase preference of a certain group. For instance, females are much more likely to buy hair masks or women's coats, while males are much more likely to buy razors or men's jackets. The comparison results are in Table 3, and comparison plots are in Figure 8; the plots are based on the specification in Model (2) with a variety of product features controlled for. We observe an interesting phenomenon: the CPF badge is most influential among demographic groups typically not reported as highly interested in sustainable consumption (i.e., older males). In contrast, for women and younger people, who were identified by past research as more motivated to buy green products (Luchs and Mooradian 2012), the effect of the CPF badge is smaller or even insignificant. This may be because these groups of consumers are already familiar with green certifications and do not rely on a third-party label for green product decisions.

Table 3. Estimation Results of Causal Impact of CPF Badge on Demand by Category

Demographic Group	ATT	p	ATT	p	RMSE	Obs. #
	Treated observations equally weighted		*Treated units equally weighted*			
Young	-0.0831	0.0005	-0.0825	0.0003	0.3195	17370
Old	-0.1652	0.0000	-0.1653	0.0000	0.3319	55777
Female	-0.0120	0.7927	-0.0043	0.9181	0.5144	292202
Male	-0.0687	0.0137	-0.0680	0.0107	0.2454	215774

Notes. For identification purposes, units whose number of untreated periods is less than 5 are dropped automatically. The categories that cater to each demographic group (Young vs. Old, Female vs. Male) are presented in Table 1.



(a) Young (b) Old (c) Female (d) Male

Figure 8. ATT of Having the CPF Badge on Demand for Different Demographic Groups: Young (a), Old (b), Female (c), and Male (d)

Notes. The gray bar denotes the number of units at the *t* period after treatment; DV is sales rank, where higher rank indicates lower demand.

6.2 CPF Badge Adoption Leads to Increased Price

In Table 4, we estimate two models to explore the causal impact of CPF on price: Model (3) includes only sales rank and rating count. Model (4) adds more variables: sales rank, rating count, review count, mean positive review scores, mean compound valence score, and various features of product descriptions (length, richness, readability) and images (colorfulness, brightness, symmetry, average face count, technical quality, and aesthetic quality). We used four augmented factors to minimize RMSE. The average treatment effect and estimated coefficients for each covariate are shown in Table 4, with time variation in the ATTs plotted in Figure 9.

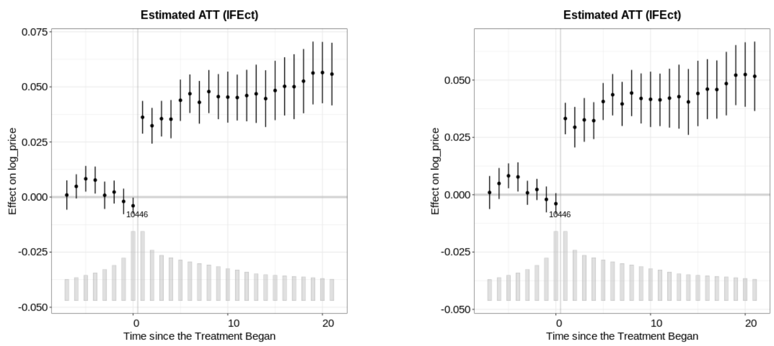
We observe a positive and significant impact of adopting the CPF badge on price, consistent even when controlling for various product features in Model (4). This aligns with previous findings that consumers are willing to pay a premium for green products (Tully and Winer 2014). Diagnostics for Model (4) in Table 4 with the price as DV, including tests for pre-trend, placebo effect, and carryover effect, are detailed in Web Appendix D.

Table 4. Estimation Results of Causal Impact of CPF Badge on Price

DV: Log (price+1)	**Model (3)**		**Model (4)**	
	ATT	p	ATT	p
Treated observations equally weighted	0.0580	0.0000	0.0535	0.0000
Treated units equally weighted	0.0492	0.0000	0.0447	0.0000
	$\hat{\beta}$	p	$\hat{\beta}$	p
Log (sales rank)	0.0128	0.0000	0.0165	0.0000
Log (rate # + 1)	-0.0074	0.0002	-0.0007	0.7504
Mean review positivity			-0.0932	0.0702
Mean rating score			0.0048	0.2956
Log (sustainability topic + 1)			-0.0210	0.0000
Log (packaging topic + 1)			-0.0276	0.0000
Description valence			0.0366	0.0081
Log (description length + 1)			0.0144	0.0873
Description richness			0.0878	0.0445
Description readability			0.0004	0.0213
Colorfulness			0.0001	0.4947
Brightness			-0.0142	0.6012
Symmetry			0.0000	0.2045
Face #			0.0456	0.0115
Image technical score			-0.0072	0.3816
Image aesthetic score			-0.0043	0.7871
RMSE	0.2043		0.2043	

Obs. #	1274958	1274958

Notes. For identification purposes, units whose number of untreated periods is less than 5 are dropped automatically.



(a) ATT in Model (3) (b) ATT in Model (4)

Figure 9. ATT of Having the CPF Badge on Price from Model (3) [a] and Model (4) [b]

Notes. The gray bar denotes number of units at the *t* period after treatment.

We also examine the average treatment effect across product categories following the same specification in Model (4), where a rich set of product image and description features are controlled for. As shown in Table 5, in five categories, the adoption of the CPF badge is associated with a positive and significant effect on price. However, for categories Clothing, Shoes & Jewelry, Office Products, and Sports & Outdoors, the effects are not significant.

Table 5. Estimation Results of Causal Impact of CPF Badge on Price by Category

Category	ATT	p	ATT	p	RMSE	Obs. #
	Treated observations equally weighted		*Treated units equally weighted*			
Beauty & Personal Care	0.0172	0.1008	0.0191	0.0574	0.1732	235846
Clothing, Shoes & Jewelry	0.0358	0.2202	0.0402	0.1640	0.0784	109238
Electronics	0.0584	0.0000	0.0578	0.0000	0.1829	203229
Grocery & Gourmet Food	0.0905	0.0000	0.0820	0.0000	0.2215	161541
Health & Household	0.0646	0.0000	0.0582	0.0000	0.2126	251479
Office Products	0.0097	0.3334	0.0131	0.2853	0.1765	151505
Sports & Outdoors	0.0421	0.1603	0.0381	0.1813	0.1563	27599
Toys & Games	0.0241	0.0971	0.0227	0.1356	0.1662	134521

Notes. For identification purposes, units whose number of untreated periods is less than 5 are dropped automatically.

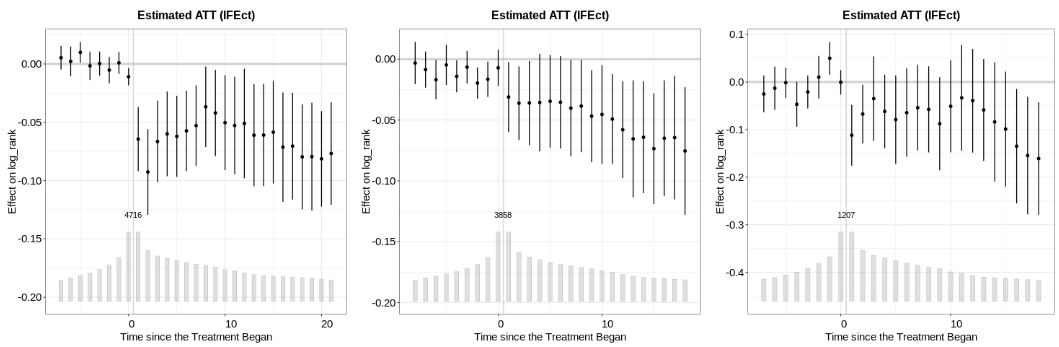
6.3 CPF Badge Adoption Leads to Decreased Market Concentration

We examine how adopting the CPF badge on one or more of a brand's products affects market concentration on Amazon. First, we analyze the impact of the badge adoption on demand for large and small brands separately. To define brand size, we use sales rank quantiles from March 1, 2023, to March 15, 2023, and divide brands within each category into three sizes—small, medium, and large—based on the 0.25 and 0.75 quantiles. We then replicate Model (2) from Table 3 for each brand size. The results show that adopting the CPF badge helps small brands become more competitive, with a more pronounced effect than on medium-sized brands (Table 6). However, the CPF badge might also increase market concentration, as the demand effect is strongest for large brands (note the different scales on the y-axis in Figure 10).

Table 6. Estimation Results of Causal Impact of CPF Badge on Demand by Brand Size

	Small Brand		**Medium Brand**		**Large Brand**	
DV: Log (sales rank)	**ATT / $\widehat{\beta}$**	**p**	**ATT / $\widehat{\beta}$**	**p**	**ATT / $\widehat{\beta}$**	**p**
Treated observations equally weighted	-0.0893	0.0000	-0.0433	0.0283	-0.1640	0.0011
Treated units equally weighted	-0.0786	0.0000	-0.0550	0.0008	-0.1557	0.0003
Log (price + 1)	0.0414	0.0909	0.1555	0.0000	0.1482	0.0000
Log (rate # + 1)	-0.3923	0.0000	-0.5456	0.0000	-0.5470	0.0000
Mean review positivity	0.2829	0.5875	0.4583	0.2236	0.1905	0.7737
Mean rating score	0.0454	0.4116	0.0552	0.2006	0.0409	0.4363
Log (sustainability topic + 1)	0.2895	0.0032	0.2966	0.0000	0.2784	0.0000
Log (packaging topic + 1)	-0.0106	0.9034	0.0546	0.2683	-0.0341	0.3497
Description compound	0.0533	0.8589	-0.1325	0.1695	-0.2217	0.0627
Log (description length + 1)	-0.2838	0.0715	-0.0924	0.0463	-0.0535	0.3809
Description richness	0.1575	0.7333	-0.6014	0.0677	-0.4232	0.0772
Description readability	-0.0041	0.2056	-0.0012	0.2300	-0.0022	0.0185
Colorfulness	-0.0004	0.8489	0.0021	0.1128	-0.0025	0.0529
Brightness	0.5150	0.4772	0.0317	0.9235	0.1243	0.7301
Symmetry	0.0000	0.3633	0.0000	0.8757	0.0000	0.3450
Face #	-0.4627	0.1708	-0.1926	0.1614	-0.2833	0.1202
Image technical score	0.2767	0.1383	0.0367	0.6440	0.1588	0.2021
Image aesthetic score	-0.1563	0.6124	-0.0317	0.7791	0.1441	0.1921
RMSE	0.2841		0.3269		0.4926	
Observations #	313102		609116		243682	

Notes. For identification purposes, units whose number of untreated periods is less than 5 are dropped automatically.



(a) ATT for Small Brands (b) ATT for Medium Brands (c) ATT for Large Brands

Figure 10. ATT of CPF Badge on Demand for Small [a], Medium [b], and Large Brands [c]

Next, we examine the relationship between CPF badge penetration (i.e., the percentage of products that have adopted the badge) and market concentration. We use the HHI (defined in section *Equilibrium Outcome*) to measure market concentration. Given the potential exit and entry of small brands, we focus only on medium- and large-sized brands, which are relatively stable, and assume that the market share held by small brands can be ignored. Additionally, we use absolute sales rank instead of relative sales rank among these medium- and large-sized brands.

To calculate the HHI, we use Equation (39). To calculate the market share, we need sales volume data. However, we have only sales rank data, and since Amazon does not disclose how it calculates sales ranks based on sales volume, we rely on the empirical transformation formula (He and Hollenbeck 2020) estimated with large datasets from Amazon for each category (not including the category of Grocery & Gourmet Food). Finally, in our data for analysis, there are 75% medium- and large-sized brands from March 15, 2023, to September 15, 2023.

We calculate the daily HHI for the market of 20 subcategories, and explore how the proportion of CPF badged products in a subcategory influences HHI with a fixed effect model:

$$HHI_{st} = b_{s0} + b_1 CPFProportion_{st} + b_2 MeanPrice_{st} + \boldsymbol{b}_3 \overrightarrow{Ratings_{st}} + \boldsymbol{b}_4 \overrightarrow{Reviews_{st}} + \boldsymbol{b}_5 \overrightarrow{Descriptions_{st}} + \boldsymbol{b}_6 \overrightarrow{Images_{st}} + \tau_t + \varepsilon_{st}, \tag{48}$$

where $CPFProportion_{st}$ denotes the proportion of CPF badged products in subcategory s at time t; and $MeanPrice_{st}$ is calculated by the mean of $log\ (price\ +\ 1)$ for products in category s at time t;

Furthermore, $Ratings_{st}$ includes *log (rate count + 1)*, mean rating score; $Reviews_{st}$ includes the mean review positivity score and occurrences of sustainability-related, packaging-, and return-related topic words; $Descriptions_{st}$ includes the mean compound valence score, length, richness, and readability of all product descriptions; and $Images_{st}$ includes the mean colorfulness, brightness, symmetry, average face count, technical quality score, and aesthetics quality scores of product images. We also control for subcategory fixed effect, b_{s0}, and time fixed effect, τ_t.

The estimation results for this fixed-effect model show a clear negative and significant effect of the CPF proportion on the HHI, indicating that as more products adopt the badge, the market becomes less concentrated (Table 7). Even though we focus on medium- and large-sized brands, assuming small brands' market share is negligible, the conclusion holds: CPF badge adoption leads to decreased market concentration, even if new small brands enter the market.

Table 7. Effect of CPF Badge Proportion on HHI

	Estimate	SD	p
CPF Badged Product Proportion	-324.3098	102.9078	0.0016
Log (price+1)	383.7037	255.4267	0.1331
Log (rate #+1)	98.6776	139.6599	0.4799
Mean review positivity	-50201.1990	5382.3200	0.0000
Mean rating score	3908.9723	350.3440	0.0000
Log (sustainability topic + 1)	1752.4851	346.7690	0.0000
Log (packaging topic + 1)	190.2718	350.2501	0.5870
Description valence	882.6418	466.1405	0.0584
Log (description length + 1)	-78.1181	149.7589	0.6020
Description richness	1332.6806	761.2951	0.0801
Description readability	-20.7897	2.4830	0.0000
Colorfulness	16.2890	1.2295	0.0000
Brightness	2593.9950	330.5784	0.0000
Symmetry	0.0040	0.0024	0.0976
Face #	-1280.8494	92.8675	0.0000
Image technical score	135.3650	122.6360	0.2698
Image aesthetic score	-1242.3769	119.7294	0.0000
(Intercept)	-410.7808	1951.6722	0.8333
Observations	3,204		
R^2	0.8592		

Notes. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

7. Discussion

We explored the impact of a unified third-party green badge on demand, price, and market concentration in an online marketplace. We developed a three-stage game theory model and found that the marketplace sets a badge threshold at one of three levels (i.e., 0, 0.5, 1) for green products. Furthermore, the theoretical model predicts that sellers set prices based on the commission rate, aiming to maximize profits, and consumers make purchase decisions, leading to profits for both the marketplace and the sellers.

Our model establishes subgame perfect equilibrium for each threshold, identifying conditions where badge adoption increases demand and prices while reducing market concentration. These conditions arise when the benefits of attracting green consumers outweigh the potential loss of non-green consumers due to higher prices. An optimal badge proportion near 50% maximizes profits, particularly when non-sustainability attributes slightly exceed those of competitors, and sustainability attributes are undervalued. Empirical data support this hypothesis, showing that a badge threshold close to 50% correlates with increased demand, higher prices, and reduced market concentration.

Next, we analyzed daily data from March 1, 2023, to September 15, 2023, for 6,606 products across 8 categories and 20 subcategories. Using computer vision and deep learning, we extracted image features such as colorfulness, brightness, and image quality; and used natural language processing techniques to analyze text features including valence, richness, and readability. We then employed multimodal machine learning models to create vector representations of images and descriptions, ensuring robust feature selection. Our data reveal that 35.39% of products had the CPF badge, supporting our hypothesis. To ensure robust causal inference, we used the IFEct estimator to address endogeneity issues and identify the CPF badge's impact on demand, price, and market concentration.

We found that the CPF badge significantly improves sales rank, indicating a positive effect on sales volume, though the impact is relatively short-lived, highlighting the importance of product imagery and descriptions. Additionally, we found that the CPF badge consistently raises prices, which is in line with previous research showing consumers' willingness to pay a premium for sustainable products.

Interestingly, the CPF badge enhances the market presence of small brands more than that of medium-sized ones; however, the badge may also increase the dominance of larger brands, ultimately reducing market concentration and fostering a more competitive marketplace. These insights are valuable for both researchers and practitioners interested in sustainable product marketing dynamics.

Our study constitutes a pioneering exploration of the broad effects of a unified eco-label—specifically, the CPF badge—on demand, pricing, and market concentration within the dynamic landscape of an e-commerce platform characterized by extensive product diversity. We elucidate the mechanisms whereby the adoption of this green badge significantly enhances demand and pricing while concurrently reducing market concentration, confirming a triple-benefit scenario supported by empirical data.

Our study provides essential insights for adjusting pricing and product strategies to maximize demand following the adoption of a green badge. Our findings are beneficial for e-commerce platforms like Amazon, enriching their understanding of the CPF badge's positive impact on expanding the customer base, notably among older and male demographics, for sustainably marketed products. Consequently, our work sets a precedent for other e-commerce and third-party online platforms considering similar eco-labeling initiatives. It offers a detailed framework for navigating sustainable marketing strategies, ensuring that the adoption of unified eco-labels is both effective in fostering competition and beneficial in enhancing consumer engagement with green products.

References

Aspara, J., X. Luo, and R. Dhar (2017) Effect of intelligence on consumers' responsiveness to a pro-environmental tax: Evidence from large-scale data on car acquisitions of male consumers, *Journal of Consumer Research*, 27(4): 448–55.

Bai, Jushan (2009), "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77(4): 1229–79.

Baker, M. J. and G. A. Churchill (1977), "The Impact of Physically Attractive Models on Advertising Evaluations," *Journal of Marketing Research*, 14(4): 538–55.

Banerjee, Subhabrata Bobby, Easwar S. Iyer, and Rajiv K. Kashyap (2003), "Corporate Environmentalism: Antecedents and Influence of Industry Type," *Journal of Marketing*, 67(2): 106–22.

Borin, N., D. C. Cerf, and R. Krishnan (2011), "Consumer effects of environmental impact in product labeling," *Journal of Consumer Marketing*, 28(1): 76–86.

Ceylan, G., K. Diehl, and D. Proserpio (2023), "Words Meet Photos: When and Why Photos Increase Review Helpfulness," *Journal of Marketing Research*.

Chen, Y. and C. Chang (2013), "Towards green trust," *Management Decision*, 51(1): 63–82.

Fairchild, Mark D. (2013), *Color Appearance Models*, Wiley.

Gobillon, Laurent and Thierry Magnac (2016), "Regional Policy Evaluation: Interactive Fixed Effects and Synthetic Controls," *Review of Economics and Statistics*, 98(3): 535–51.

Gorn, Gerald J, Amitava Chattopadhyay, Tracey Yi, and Darren W Dahl (1997), "Effects of Color as an Executional Cue in Advertising: They're in the Shade," *Management Science*, 43 (10), 1387–1400.

He, S., B. Hollenbeck, D. Proserpio (2022), "The Market for Fake Reviews," *Mar. Sci.*, 41 (5): 896–921.

Johnstone, Micael-Lee and Lay Peng Tan (2015), "Exploring the Gap Between Consumers' Green Rhetoric and Purchasing Behaviour," *Journal of Business Ethics*, 132(2): 311–28.

Kelly, W. A. (1981), "A Generalized Interpretation of the Herfindahl Index," *Sou. Econ. J.*, 48(1): 50.

Kreitler, Hans and Shulamith Kreitler (1972), "The Model OF Cognitive Orientation: Towards a Theory of Human Behavior," *British Journal of Psychology*, 63(1): 9–30.

Liu, Licheng, Ye Wang, and Yiqing Xu (2022), "A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data," *American Journal of Political Science*.

Liu, X., D. Lee, and K. Srinivasan (2019), "Large-Scale Cross-Category Analysis of Consumer Review Content on Sales Conversion Leveraging Deep Learning," *Journal of Marketing Research*, 56(6): 918–43.

Luchs, M. G, R. W. Naylor, J. R Irwin, and R. Raghunathan (2010), "The Sustainability Liability: Potential Negative Effects of Ethicality on Product Preference," *Journal of Marketing*, 74(5): 18–31.

Luchs, Michael G. and Todd A. Mooradian (2012), "Sex, Personality, and Sustainable Consumer Behaviour: Elucidating the Gender Effect," *Journal of Consumer Policy*, 35 (1), 127–44.

Mai, Long, Hailin Jin, and Feng Liu (2016), "Composition-Preserving Deep Photo Aesthetics Assessment," in *2016 CVPR*, IEEE, 497–506.

Mazar, N. and C. Zhong (2010), "Do Green Products Make Us Better People?," *Psy. Sci.*, 21(4): 494–98.

Menon, Ajay and Anil Menon (1997), "Enviropreneurial Marketing Strategy: The Emergence of Corporate Environmentalism as Market Strategy," *Journal of Marketing*, 61(1): 51.

M., Patrick, N. Kangun, and W. B. Locander (1978), "Environmentally Concerned Consumers-Racial Variations," *Journal of Marketing*, 42(4): 61–66.

Murray, N., L. Marchesotti, and F. Perronnin (2012), "AVA: A large-scale database for aesthetic visual analysis," in *2012 CVPR*, IEEE, 2408–15.

Newman, George E., Margarita Gorlin, and Ravi Dhar (2014), "When Going Green Backfires: How Firm Intentions Shape the Evaluation of Socially Beneficial Product Enhancements," *Journal of Consumer Research*, 41(3): 823–39.

Olsen, Mitchell C, Rebecca J Slotegraaf, and Sandeep R Chandukala (2014), "Green Claims and Message Frames: How Green New Products Change Brand Attitude," *Journal of Marketing*, 78 (5), 119–37.

Parguel, B., F. Benoît-Moreau, and F. Larceneux (2011), "How Sustainability Ratings Might Deter 'Greenwashing': A Closer Look at Ethical Corporate Communication," *Journal of Business Ethics*, 102(1): 15–28.

Park, Sungsik, Man Xie, and Jinhong Xie (2023), "Frontiers: Framing Price Increase as Discount: A New Manipulation of Reference Price," *Marketing Science*, 42 (1), 37–47.

Peattie, K., and S. Peattie (2009) Social marketing: A pathway to consumption reduction? *Journal of Business Research*, 62(2): 260–68.

Pieters, Rik, Michel Wedel, and Rajeev Batra (2010), "The Stopping Power of Advertising: Measures and Effects of Visual Complexity," *Journal of Marketing*, 74 (5), 48–60.

Radford, Alec, Jong Wook Kim, Chris Hallacy, et al. (2021), "Learning Transferable Visual Models From Natural Language Supervision."

Reczek, R. W., J. R Irwin, D. M Zane, and K. R Ehrich (2018), "That's Not How I Remember It: Willfully Ignorant Memory for Ethical Product Attribute Information," *Journal of Consumer Research*, 45(1): 185–207.

Reimann, Martin, Judith Zaichkowsky, Carolin Neuhaus, et al. (2010), "Aesthetic package design: A behavioral, neural, and psychological investigation," *Journal of Consumer Psychology*, 20(4): 431–41.

Sen, Sankar and C.B. Bhattacharya (2001), "Does Doing Good Always Lead to Doing Better? Consumer Reactions to Corporate Social Responsibility," *Journal of Marketing Research*, 38(2): 225–43.

Singh, Nirvikar and Xavier Vives (1984), "Price and Quantity Competition in a Differentiated Duopoly," *RAND Journal of Economics*, 15(4): 546–54.

Sussman, Abigail B. and Rourke L. O'Brien (2016), "Knowing When to Spend: Unintended Financial Consequences of Earmarking to Encourage Savings," *Journal of Marketing Research*, 53(5): 790–803.

Talebi, Hossein and P Milanfar (2018), "NIMA: Neural Image Assessment," *IEEE Transactions on Image Processing*, 27, 3998–4011.

Townsend, Claudia and Suzanne B. Shu (2010), "When and how aesthetics influences financial decisions," *Journal of Consumer Psychology*, 20 (4), 452–58.

Tully, Stephanie M. and Russell S. Winer (2014), "The Role of the Beneficiary in Willingness to Pay for Socially Responsible Products: A Meta-analysis," *Journal of Retailing*, 90 (2), 255–74.

White, K., J. Hardisty, and R. Habib (2019), "The Elusive Green Consumer," *Harvard Bus. Rev*.

Xu, Yiqing (2017), "Generalized Synthetic Control Method: Causal Inference with Interactive Fixed Effects Models," *Political Analysis*, 25 (1), 57–76.

Zhang, S., D. Lee, P. V. Singh, and K. Srinivasan (2022), "What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features," *Management Science*, 68(8): 5644–66.

Zhou, B. and T. Zou (2023), "Competing for Recommendations: The Strategic Impact of Personalized Product Recommendations in Online Marketplaces," *Marketing Science*, 42(2): 360–76.

Web Appendix: Sustainability and Competition on Amazon

Web Appendix A. Green Certificates

We collected daily data on the product detail page from March 1st to September 15th, 2023, covering 14,000 products distributed across eight categories and 20 sub-categories. These categories were selected based on their high proportion of green products on Amazon. Table W1 presents the assortment of products categorized by type, sub-type, and the prevalence of associated green certifications. We further create two variables based on the primary age and gender demographic of the consumer base of the subcategory products: the 'Age Group' column classifies 'young' or 'old', while the 'Gender Group' column denotes 'female' or 'male.' We summarize the sample statistics by the group variables. In addition, the term 'all' is employed to indicate that the respective sub-category is universally applicable, transcending age and gender distinctions, thus suitable for both males and females, and for both younger and older consumers.

Table W1. Number of Units and Main Certificate for Each Sub-Category

Category	Sub-Category	Age Group	Gender Group	Reference	Total	Treated	*Main Certificate*
Beauty & Personal Care	black hair dye	old	all	(Clarke and Korotchenko 2010)	359	67	*USDA Organic*
	razor	all	male	(Chang and Lipner 2021)	363	82	
	hair mask	young	female	(Lai 2022)	500	195	*Cradle to Cradle Certified*
Clothing, Shoes & Jewelry	woman coat	all	female		272	11	*Bluesign*
	man jacket	all	male		294	7	
Electronics	speaker	old	all	(NIDCD 2023)	579	357	*Carbon Neutral Certified by SCS Global Services*
	wireless earbuds	young	all	(Nguyen 2020)	474	259	
Grocery & Gourmet Food	tea	old	all	(Ruxton et al. 2021)	417	249	*USDA Organic*
	energy drink	young	all	(Levitt 2023)	420	179	
Health & Household	manual toothbrush	old	all	(Vogels 2019)	227	37	*The Forest Stewardship Council*
	electronic toothbrush	young	all	(Vogels 2019)	355	71	*Climate neutral by ClimatePartner*

	Aromatherapy	all	female	(Yan et al. 2019)	344	166	*USDA Organic*
	lighter	all	male	(Waldron 1991)	377	103	*Compact by Design (Amazon)*
Office Products	ink	all	all		372	153	*Climate neutral by ClimatePartner*
	pen refill	all	all		413	177	
Sports & Outdoors	yoga equipment	all	female	(Gregoire 2013)	93	33	*Climate neutral by ClimatePartner*
	biking equipment	all	male	(Lindsey 2019)	50	34	
Toys & Games	card games	old	all	(Blocker, Wright, and Boot 2014)	358	114	*The Forest Stewardship Council*
	video games	young	male	(Orlando 2018)	34	10	
	dolls	all	female	(Kimont 2017)	305	34	

Notes. "Age Group" and "Gender Group" columns signify the likely purchasers based on age and gender. The "Main Certificate" column lists the dominant green certificate, recognizing that each sub-category may have multiple types of green certificates.

Web Appendix B. Data Visualization

During our six-and-a-half months of data, we noted treatment reversals where products either gained or lost their badge status over time. To better illustrate treatment reversal over time, we plot out the history of CPF badge status for a random sample of 100 distinct products (Figure W1). We can observe that for some of the products, the treatment status has changed multiple times, which is very likely due to Amazon adjusting the badge threshold. Therefore, when choosing the causal inference method, to use most of the data available, we will need a method that allows for treatment reversal. The y-axis represents the product ID, while the x-axis represents the day. Each dot on the plot represents whether the corresponding product (product ID on the y-axis) is badged (dark blue) or not badged (light blue) on the focal day (date on the x-axis).

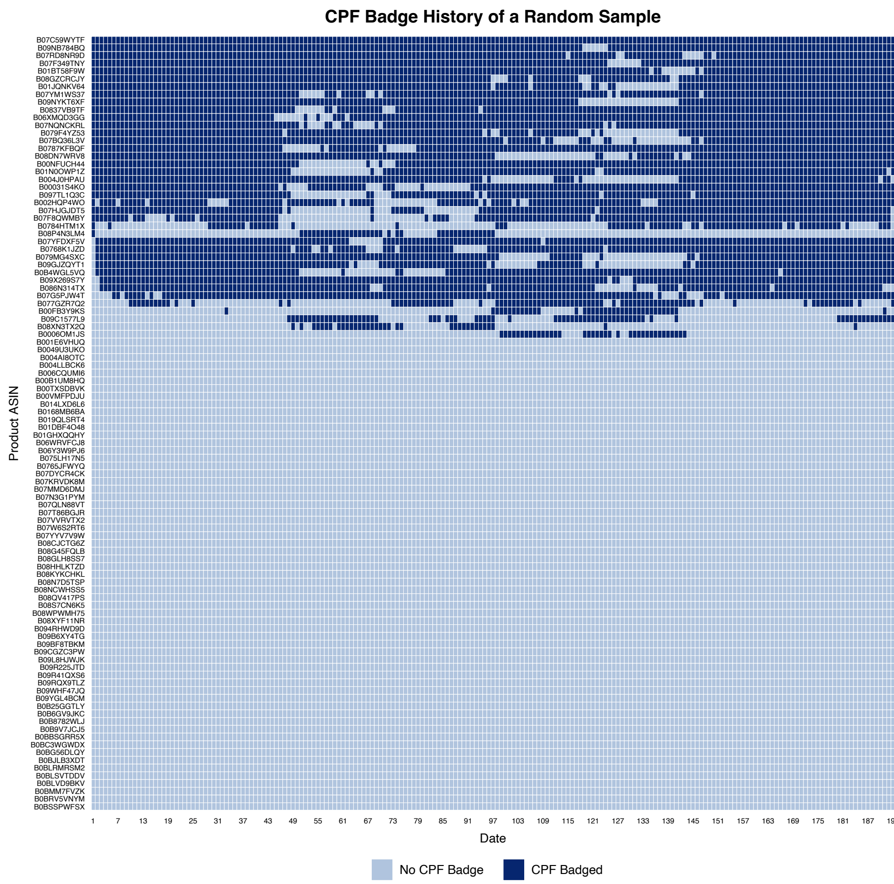


Figure W1. Histories of Treatment Status of a Random Sample

Web Appendix C. Technical Details on IFECT Estimator

3.1 Counterfactual Estimator

We begin by introducing the fixed effects counterfactual (FEct) estimator in its basic form, utilizing the linear expressions for $f(\cdot)$ and $h(\cdot)$ as outlined in Assumption 1, based on the identification strategy described earlier. Suppose that

$$Y_{it}(0) = \boldsymbol{X'}_{it}\beta + \alpha_i + \tau_t + \varepsilon_{it}, \tag{W1}$$

where $Y_{it}(0)$ denoting the outcome when not being treated; $\boldsymbol{X'}_{it}$ is a vector of covariates; α_i is unit fixed effect while τ_t is time fixed effect; ε_{it} is the idiosyncratic error term.

This is very similar to two-way fixed effect (TWFE) model, but differently we impose a constraint

$$\sum_{D_{it}=0} \alpha_i = \sum_{D_{it}=0} \tau_t, \quad \text{(W2)}$$

which ensures identification and makes the grand mean (a common intercept across units and time) redundant. To be more specific, Arkhangelsky and Imbens (2021) show that with the additive unit and time FE, any estimator that aims at identifying a convex combination of δ_{it} can be written as a weighted average of Y_{it}, where the weights $\{w_{it}\}_{1\le i\le N, 1\le t\le T}$ must satisfy following conditions:

1. $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} w_{it}D_{it} = 1$
2. $\sum_{t=1}^{T} w_{it} = 0$ for any i
3. $\sum_{i=1}^{N} w_{it} = 0$ for any t
4. $w_{it}D_{it} \ge 0$ for any pair of (i,t)

Weights from a conventional TWFE model meet conditions 1 ~ 3. However, the last condition is violated which might lead to biases from improper weighting. In comparison, with the newly imposed constraint, condition 4 is also satisfied, as FEct also imposes $w_{it:D_{it}=1} = 1/|\mathcal{M}|$, where $\mathcal{M}$ is a set of observations under the treatment condition. We denote $\mathcal{O} = \{(i,t)|D_{it} = 0\}$ as the set of observations under control condition. In this way, we rewrite the FEct as a weighting estimator where each treated observation is matched with its predicted counterfactual $\hat{Y}_{it}(0) = \mathbf{W}'^{(\mathrm{it})}\mathbf{Y}_{\mathcal{O}}$, which is the weighted sum of untreated observations and weights $\mathbf{W}'^{(\mathrm{it})} = (\dots, W_{js}^{(it)}, \dots)$, $(j,s) \in \mathcal{O}$ satisfy

$$\sum_{s:(i,s)\in\mathcal{O}} W_{is}^{(it)} = \sum_{j:(j,t)\in\mathcal{O}} W_{jt}^{(it)} = 1, \sum_{j:s\neq t,(j,s)\in\mathcal{O}} W_{js}^{(it)} = \sum_{s:j\neq i,(j,s)\in\mathcal{O}} W_{js}^{(it)} = 0. \quad \text{(W3)}$$

The advantages of such an estimator include (1) Because treated observations of early treatment adopters never serve as controls for late treatment adopters, we prevent the negative weights problem[8] from its root cause; (2) Compared with difference-in-difference (DID), this method is more efficient because it uses most available data without imposing stronger functional

[8] The issue of "negative weights" emerges when units that have already received treatment are utilized as controls, leading to the subtraction of changes in their outcomes, which may encompass time-varying treatment effects. While this doesn't indicate a flaw in the design regarding non-parallel trends in counterfactual outcomes, it underscores the need for caution when employing TWFE estimators to summarize treatment effects (de Chaisemartin and D'Haultfœuille 2020; Goodman-Bacon 2021).

form assumptions; (3) Comparison within each matched pair removes the biases caused by improper weighting that plague conventional FE models.

3.2 Basic Assumptions

We make three basic assumptions to ensure the identification of the treatment effect, similar to those made by Liu, Wang, and Xu (2022).

Assumption 1. We do not consider temporal and spatial interaction and only assume a parametric functional form on the covariates:

$$Y_{it}(0) = f(\boldsymbol{X}_{it}) + h(\boldsymbol{U}_{it}) + \varepsilon_{it}, \tag{W4}$$

where $f(\cdot)$ and $h(\cdot)$ are known, parametric functions.

Assumption 2. We assume strict exogeneity between outcome variable and covariates:

$$\varepsilon_{it} \perp \{D_{js}, \boldsymbol{X}_{js}, \boldsymbol{U}_{js}\}, \forall i, j \in \{1,2, \ldots, N\}, \forall t, s \in \{1,2, \ldots, T\}.$$

Assumption 3. There exists a low-dimensional decomposition of $h(\boldsymbol{U}) = \boldsymbol{\Lambda}\mathbf{F}$, where $\boldsymbol{\Lambda} \in \mathbb{R}^{N\times r}$ is a matrix of factor loadings and $\mathbf{F} \in \mathbb{R}^{r\times T}$ is a matrix of factors, with the dimension $r \ll \min\{N, T\}$.

Assumptions 1 and 2 together exclude potential anticipation effects (Wang 2021) which leads to under-identification of the causal effects (Borusyak, Jaravel, and Spiess 2021); feedback loop from Y_{it-1} to D_{it}, and lagged dependent variables from Y_{it-1} to from Y_{it} are also ruled out; besides, the treatment effect of D_{it} on Y_{it} is separable from the influences of $\boldsymbol{X}_{it}$ and $\boldsymbol{U}_{it}$. When unobserved confounders $\boldsymbol{U}_{it}$ exist, treatment assignment is dependent on observed untreated outcomes, thus, we are operating under a special case of missing not at random (Rubin 1976). Assumption 3 allows us to break this dependency by controlling for $\boldsymbol{U}_{it}$ approximated using data.

3.3 Algorithm Estimation

The basic logic of estimation includes 4 steps. We denote $\mathcal{O} = \{(i,t)|D_{it} = 0\}$ as the set of observations under control condition, and $\mathcal{M} = \{(i,t)|D_{it} = 1, C_i = 1\}$ as the set of observations under treatment condition which will only be among those treated units. In the first step, based on Assumption 1 and 3, we fit a model of the response surface Y_{it} on the subsect of untreated observations $\mathcal{O}$, which allow us to obtain estimated $\hat{f}$ and $\hat{h}$. In the second step, we predict the counterfactual outcome $Y_{it}(0)$ for each treated observation based on $\hat{Y}_{it}(0) = \hat{f}(\boldsymbol{X}_{it}) + \hat{h}(\boldsymbol{U}_{it})$ for all $(i,t) \in \mathcal{M}$. For the third step, we estimate the individualistic treatment effects δ_{it} based on $\hat{\delta}_{it} = Y_{it} - \hat{Y}_{it}(0)$ for all treated observations $(i,t) \in \mathcal{M}$. Finally, we can calculate ATT using predicted $\hat{\delta}_{it}$:

$$\widehat{ATT} = \frac{1}{|\mathcal{M}|} \sum_{\mathcal{M}} \hat{\delta}_{it}. \tag{W5}$$

For the IFEct estimator, we assume in the iteration round h we have $\hat{\mu}^{(h)}$, $\hat{\alpha}_i^{(h)}$, $\hat{\tau}_t^{(h)}$, $\hat{\lambda}_i^{(h)}$, $\hat{f}_t^{(h)}$ and $\hat{\beta}^{(h)}$. Note that $\hat{\mu}^{(h)} = \boldsymbol{X'}_{it}\hat{\beta}^{(h)}$ where $\boldsymbol{X'}_{it}$ is a vector of covariates. For the untreated, suppose that we have

$$\dot{Y}_{it}^{(h)} \coloneqq Y_{it} - \hat{\mu}^{(h)} - \hat{\alpha}_i^{(h)} - \hat{\tau}_t^{(h)} - \hat{\lambda}'^{(h)}_i \hat{f}_t^{(h)}. \tag{W6}$$

Now, in round h+1, we update $\hat{\beta}^{(h+1)}$ with untreated data only

$$\hat{\beta}^{(h+1)} = \left(\sum_{D_{it}=0} \boldsymbol{X}_{it} \boldsymbol{X'}_{it}\right)^{-1} \sum_{D_{it}=0} \boldsymbol{X}_{it} \dot{Y}_{it}^{(h)}, \tag{W7}$$

where $\left(\sum_{D_{it}=0} \boldsymbol{X}_{it} \boldsymbol{X'}_{it}\right)^{-1}$ is fixed and is not updated in every round.

For all observations, define

$$W_{it}^{(h+1)} \coloneqq \begin{cases} Y_{it} - \boldsymbol{X'}_{it}\hat{\beta}^{(h+1)}, D_{it} = 0, \\ \hat{\mu}^{(h)} + \hat{\alpha}_i^{(h)} + \hat{\tau}_t^{(h)} + \hat{\lambda}'^{(h)}_i \hat{f}_t^{(h)}, D_{it} = 1, \end{cases} \tag{W8}$$

based on which for all untreated observations we calculate $W_{it}^{(h)}$, while for all treated observations calculate conditional expectation:

$$\mathbb{E}\left(W_{it}^{(h+1)} | \hat{\lambda}_i^{(h)}, \hat{f}_t^{(h)}\right) = \hat{\mu}^{(h)} + \hat{\alpha}_i^{(h)} + \hat{\tau}_t^{(h)} + \hat{\lambda}'^{(h)}_i \hat{f}_t^{(h)}. \tag{W9}$$

We denote the weights as

$$W_{\cdot\cdot}^{(h+1)} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it}^{(h+1)}, \tag{W10}$$

$$W_{i\cdot}^{(h+1)} = \frac{1}{T} \sum_{t=1}^{T} W_{it}^{(h+1)}, \forall i \in \{1,2, \dots, N\} \tag{W11}$$

$$W_{\cdot t}^{(h+1)} = \frac{1}{N}\sum_{i=1}^{N} W_{it}^{(h+1)}, \forall t \in \{1,2,\dots,T\} \quad \text{(W12)}$$

$$\widetilde{W}_{it}^{(h+1)} = W_{it}^{(h+1)} - W_{i\cdot}^{(h+1)} - W_{\cdot t}^{(h+1)} + W_{\cdot\cdot}^{(h+1)}, \quad \text{(W13)}$$

where we impose restrictions: $\sum_{i=1}^{N} \alpha_i = \sum_{t=1}^{T} \tau_t = \sum_{i=1}^{N} \lambda_i = \sum_{t=1}^{T} f_t = 0$.

Now, we update estimates of factors and factor loadings by minimizing the least squares objective function with $\mathbf{W}^{(h+1)} = \left[\widetilde{W}_{it}^{(h+1)}\right]_{\forall i,t}$ based on Assumption 3 with two restrictions to ensure the identification of matrix $\tilde{\mathbf{F}}$ and $\widetilde{\boldsymbol{\Lambda}}$:

$\left(\hat{\mathbf{F}}^{(h+1)}, \widehat{\boldsymbol{\Lambda}}^{(h+1)}\right) = \underset{(\tilde{\mathbf{F}},\widetilde{\boldsymbol{\Lambda}})}{\operatorname{argmin}} \ \mathbf{tr}\left[(\mathbf{W}^{(h+1)} - \tilde{\mathbf{F}}\widetilde{\boldsymbol{\Lambda}}')'(\mathbf{W}^{(h+1)} - \tilde{\mathbf{F}}\widetilde{\boldsymbol{\Lambda}}')\right]$, s.t., $\frac{\tilde{\mathbf{F}}'\tilde{\mathbf{F}}}{T} = \mathbf{I}_r$, $\widetilde{\boldsymbol{\Lambda}}'\widetilde{\boldsymbol{\Lambda}}$ is diagonal. Throughout, for a vector or matrix A, its norm is defined as $\|A\| = (\text{tr}(A'A))^{1/2}$.

Based on all these, we update estimates of grand mean and two-way fixed effects:

$$\hat{\mu}^{(h+1)} = W_{\cdot\cdot}^{(h+1)}, \hat{\alpha}_i^{(h)} = W_{i\cdot}^{(h+1)} - W_{\cdot\cdot}^{(h+1)}, \hat{\tau}_t^{(h)} = W_{\cdot t}^{(h+1)} - W_{\cdot\cdot}^{(h+1)}, \quad \text{(W14)}$$

based on which we can finally estimate

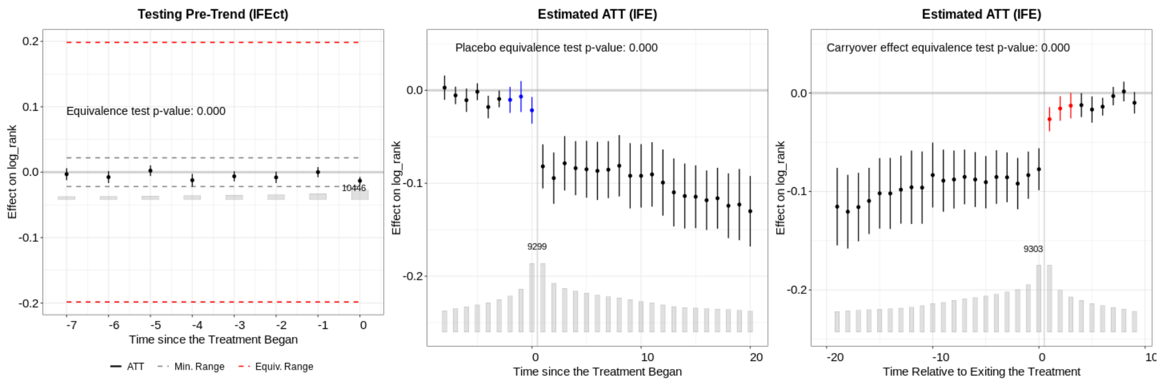
$$\hat{Y}_{it}(0) = \boldsymbol{X'}_{it}\hat{\beta} + \hat{\alpha}_i + \hat{\tau}_t + \hat{\lambda}_i'\hat{f}_t, \forall i,t\ D_{it} = 1. \quad \text{(W15)}$$

The last step is simply to estimate ATT.

Web Appendix D. More Explorations on the Impact of CPF Badge

4.1 Robustness Check of CPF Badge's Impact on Demand

For Model (2) in the main document (Table 3), we run diagnostics including a test for pre-trend, a test for placebo effect, and a test for carryover effect, as illustrated below. We use two-one-sided t (TOST) test, a form of equivalence test (Schuirmann 1987). This differs from traditional hypothesis testing which typically focuses on finding evidence for significant differences rather than similarity or equivalence. The TOST examines whether the 90% confidence intervals for estimated Average Treatment Effects (ATTs) in the pre-treatment or post-treatment periods exceed a predefined range, termed the equivalence range. A smaller p-value from the TOST indicates a better fit for the current model. We can see that Model (2) passes the three tests, and there's obviously no pre-trend (Figure W2 [a]), no placebo effect in the pre-treatment period (Figure W2 [b]), and no carryover effect in the post-treatment period (Figure W2 [c]).



(a) a test for pre-trend (b) a test for placebo effect (c) a test for carryover effect

Figure W2. Diagnostic Tests for Model (2)

Notes. The grey bar denotes the number of units at the *t* period after treatment. DV is sales rank: higher rank indicates lower demand.

As a robustness check, we control for vector representation from multi-modal model, CLIP, as described in subsection *Multi-Modal Vector Representations*. Consistently, a positive and significant effect of having a badge on demand is found (Table W2).

Table W2. Estimation Results of Causal Impact of CPF Badge on Demand – Robustness Check with CLIP Features

DV: Log (sales rank)	**Clip Feature-Based Model**	
	ATT	p
Treated observations equally weighted	-0.1094	0.0000
Treated units equally weighted	-0.1051	0.0000
	$\hat{\beta}$	p
Log (price+1)	0.1262	0.0000
Log (rate #+1)	-0.5216	0.0000
Mean review positivity	0.4320	0.0884

Mean rating score	0.0480	0.0514
Log (sustainability topic+1)	0.2810	0.0000
Log (packaging topic+1)	0.0033	0.9220
Image Vector Dim-0	0.0307	0.1809
Image Vector Dim-1	-0.0255	0.3235
Image Vector Dim-2	-0.0515	0.0420
Image Vector Dim-3	0.0265	0.1797
Image Vector Dim-4	-0.0117	0.7817
Image Vector Dim-5	-0.0862	0.0036
Image Vector Dim-6	0.0483	0.1657
Image Vector Dim-7	-0.0693	0.1230
Image Vector Dim-8	-0.1671	0.0000
Image Vector Dim-9	-0.0110	0.7870
Description Vector Dim-0	-0.0216	0.4481
Description Vector Dim-1	-0.0503	0.5824
Description Vector Dim-2	-0.0531	0.2611
Description Vector Dim-3	-0.0976	0.2506
Description Vector Dim-4	-0.0397	0.5056
Description Vector Dim-5	-0.0175	0.7737
Description Vector Dim-6	0.0989	0.0753
Description Vector Dim-7	0.0235	0.7963
Description Vector Dim-8	-0.0005	0.9951
Description Vector Dim-9	-0.0325	0.6922
RMSE	0.3460	
Obs. #	1274958	

Notes. For identification purposes, units whose number of untreated periods <5 are dropped automatically.

Moreover, to obtain a cleaner inference and as an alternative way of robustness check and sanity check, we only look at products with price, image, and description unchanged during the 3-month period from May 31st to Aug 30th 2023. Consistently, positive and significant effect of adopting a badge on demand is found (Table W3).

Table W3. Estimation Results of Causal Impact of CPF Badge on Demand

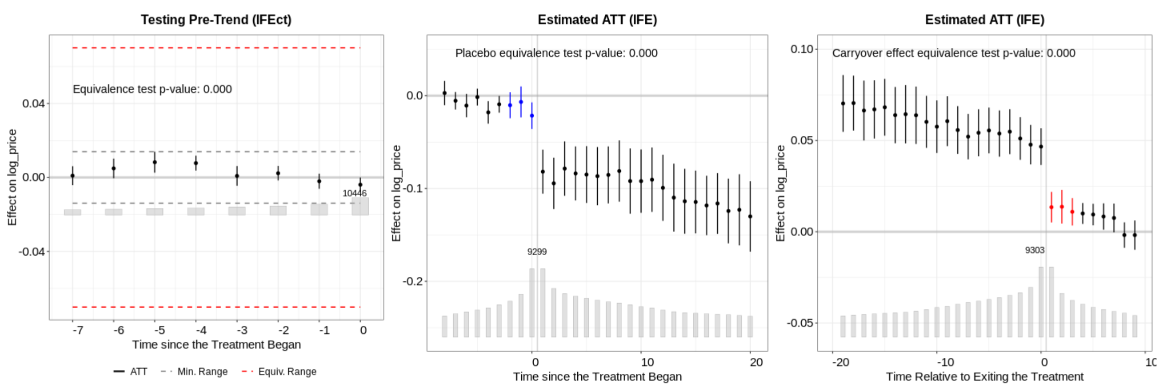
DV: Log (sales rank)	**Clean Sample Model**	
	ATT	p
Treated observations equally weighted	-0.1420	0.0013
	$\hat{\beta}$	p
Log (rate #+1)	-0.4766	0.0574

Mean review positivity	-9.3676	0.4453
Mean rating score	3.1771	0.0829
Log (sustainability topic+1)	0.1023	0.8077
Log (packaging topic+1)	-1.5230	0.2560
RMSE	0.1280	
Obs. #	43425	

Notes. For identification purposes, units whose number of untreated periods <5 are dropped automatically.

4.2 Robustness Check of CPF Badge's Impact on Price

For Model (4) in the main document (Table 6), we run diagnostics including a test for pre-trend, a test for placebo effect, and a test for carryover effect, as illustrated below. We use TOST test, same as in the previous subsection. We can see that Model (4) passes the three tests, and there's obviously no pre-trend (Figure W3 [a]), no placebo effect in the pre-treatment period (Figure W3 [b]), and no carryover effect in the post-treatment period (Figure W3 [c]).



(a) a test for pre-trend (b) a test for placebo effect (c) a test for carryover effect

Figure W3. Diagnostic Tests for Model (4)

Notes. The grey bar denotes the number of units at the *t* period after treatment.

As a robustness check, we control for vector representation from multi-modal model, CLIP, as described in subsection *Multi-Modal Vector Representations*. Consistently, a positive and significant effect of having a badge on price is found (Table W4).

Table W4. Estimation Results of Causal Impact of CPF Badge on Price – Robustness Check with CLIP Features

DV: Log (sales rank)	**Clip Feature-Based Model**	
	ATT	p
Treated observations equally weighted	0.0544	0.0000
Treated units equally weighted	0.0459	0.0000
	$\hat{\beta}$	p
Log (sales rank)	0.0159	0.0000
Log (rate #+1)	-0.0012	0.6013
Mean review positivity	-0.0920	0.0600
Mean rating score	0.0038	0.4303
Log (sustainability topic+1)	-0.0212	0.0000
Log (packaging topic+1)	-0.0279	0.0000
Image Vector Dim-0	-0.0008	0.8826
Image Vector Dim-1	-0.0036	0.5386
Image Vector Dim-2	0.0061	0.2767
Image Vector Dim-3	-0.0043	0.4078
Image Vector Dim-4	-0.0132	0.1880
Image Vector Dim-5	0.0078	0.2188
Image Vector Dim-6	-0.0019	0.7116
Image Vector Dim-7	0.0074	0.4453
Image Vector Dim-8	0.0085	0.3442
Image Vector Dim-9	-0.0077	0.4602
Description Vector Dim-0	0.0012	0.7496
Description Vector Dim-1	0.0085	0.4737
Description Vector Dim-2	0.0163	0.2038
Description Vector Dim-3	0.0266	0.0065
Description Vector Dim-4	-0.0106	0.4354
Description Vector Dim-5	-0.0055	0.7002
Description Vector Dim-6	0.0239	0.0127
Description Vector Dim-7	0.0160	0.2828
Description Vector Dim-8	-0.0170	0.0835
Description Vector Dim-9	-0.0061	0.6147
RMSE	0.2043	
Obs. #	1274958	

Notes. For identification purposes, units whose number of untreated periods <5 are dropped automatically.

4.3 CPF Badge and Search Rank

To eliminate the concerns on major influence from search engine, we additionally collected daily data on the same 14K products from Dec 25th 2023 to Jan 27th 2024; we also collected the search rank for each of the 20 product sub-categories on a daily basis. After removing missing observation on best seller ranks, price, rate score, and rate count, we've got 142 distinct products that have the badge status changed during this one-month and has appeared at least once in the search results. Amazon only show parts of the products in the search results based on the search ranking algorithm which change every day, and for those products not appearing in the search results we give a large value of 1000 (the maximum search rank is 650). In this way, we obtained an unbalanced panel without interpolation of 142 distinct products across 35 days.

Then, we use a two-way fixed effect model to explore whether search rank is affected by whether having a badge, controlling for product fixed effect and time fixed effect:

$$Log(SearchRank_{jt}) = \alpha_j + \tau_t + b_1 CPFBadged_{jt} + b_2 Log(SalesRank_{jt})$$
$$+b_2 Log(Price_{jt} + 1) + b_3 Log(Rate\#_{jt} + 1) + b_4 RateScore_{jt} + \varepsilon_{jt}, \quad \text{(W16)}$$

where $SearchRank_{jt}$ is the ranking of product j on day t in the search result for sub-category it belongs to. α_j is the product fixed effect and τ_t captures time fixed effect. $CPFBadged_{jt}$ is a dummy variable indicating whether product j is badged on day t. $Price_{jt}$, $SalesRank_{jt}$, $Rate\#_{jt}$, and $RateScore_{jt}$ are the time-varying price, sales rank of the focal category (as a proxy for demand), rating count, and rating score on day t for product j. ε_{jt} is the error term. The estimation results are shown below (Table W5), where we can obviously see that whether having a badge has no significant effect on search rank, alleviating concerns to some extent that CPF badge has a positive and significant effect on sales rank because of its impact on the search rank.

Table W5. Estimation Results of Effect of CPF Badge on Search Rank

VARIABLES	DV: Log(Search Rank)
CPF Badged	-0.0202
	(0.0183)

Log(Price + 1)	0.00207
	(0.00634)
Log(Sales Rank)	0.00912*
	(0.00498)
Log(Rate # + 1)	0.00157
	(0.00543)
Rate Score	-0.0253
	(0.0286)
Observations	3,473
Number of Products	142

Notes. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

References

Arkhangelsky, Dmitry and Guido Imbens (2021), "Double-Robust Identification for Causal Panel Data Models," Cambridge, MA.

Blocker, K.A., T.J. Wright, and W.R. Boot (2014), "Gaming preferences of aging generations," *Gerontechnology*, 12 (3).

Borusyak, Kirill, Xavier Jaravel, and Jann Spiess (2021), "Revisiting Event Study Designs: Robust and Efficient Estimation."

de Chaisemartin, Clément and Xavier D'Haultfœuille (2020), "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects," *American Economic Review*, 110 (9), 2964–96.

Chang, Michelle J. and Shari R. Lipner (2021), "Gender-based stereotyping and cost discrepancies for razors," *International Journal of Women's Dermatology*, 7 (4), 494–95.

Clarke, Laura Hurd and Alexandar Korotchenko (2010), "Shades of grey: to dye or not to dye one's hair in later life," *Ageing and Society*, 30 (6), 1011–26.

Goodman-Bacon, Andrew (2021), "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 225 (2), 254–77.

Gregoire, Carolyn (2013), " YOGA WOMEN WASHINGTON POST The Real Reason Yoga Is Still Dominated By Women," *HUFFPOST*.

Kimont, Kaitlin (2017), "Science May Explain Why Little Girls Like Dolls," *romper*.

Lai, Laura (2022), “TikTok hair care and scalp care trends – how is it evolving the beauty landscape and influencing Generation Z?,” *CRODA*.

Levitt, Aimee (2023), “Prime targets: why are teenagers so obsessed with energy drinks?,” *The Guardian*.

Lindsey, Greg (2019), “Bicycles, Gender, and Risk: Driver Behaviors When Passing Cyclists.”

Liu, Licheng, Ye Wang, and Yiqing Xu (2022), “A Practical Guide to Counterfactual Estimators for Causal Inference with Time-Series Cross-Sectional Data,” *American Journal of Political Science*.

Nguyen, Dexter (2020), “Analyzing Customer Satisfaction of Apple AirPods Using Exploratory Data Analysis and Classification Techniques (Part 2),” *Towards Data Science*.

NIDCD (2023), “Age-Related Hearing Loss (Presbycusis).”

Orlando, Joanne (2018), “Why men and boys crave their video games,” *The Sydney Morning Herald*.

Rubin, Donald B. (1976), “Inference and Missing Data,” *Biometrika*, 63 (3), 581.

Ruxton, Carrie, Gill Jenkins, Robin Seymour, and Vanessa Facenda (2021), “Drink tea to good health across the ages,” *Tea & Coffee Trade Journal*.

Schuirmann, Donald J. (1987), “A comparison of the Two One-Sided Tests Procedure and the Power Approach for assessing the equivalence of average bioavailability,” *Journal of Pharmacokinetics and Biopharmaceutics*, 15 (6), 657–80.

Vogels, Emily A. (2019), “Millennials stand out for their technology use, but older generations also embrace digital life,” *Pew Research Center*.

Waldron, Ingrid (1991), “Patterns and causes of gender differences in smoking,” *Social Science & Medicine*, 32 (9), 989–1005.

Wang, Ye (2021), “Causal Inference under Temporal and Spatial Interference.”

Yan, Min-Ren, Chun-Han Wang, Nelson Jose Cruz Flores, and Ying-Yueh Su (2019), “Targeting Open Market with Strategic Business Innovations: A Case Study of Growth Dynamics in Essential Oil and Aromatherapy Industry,” *Journal of Open Innovation: Technology, Market, and Complexity*, 5 (1), 7.