Signaling under Bilateral Uncertainty: *Do Green Consumers Lead to More Greenwashing?*[*]

Dolores Garrido[†], Ana Espinola-Arredondo[‡], and Felix Munoz-Garcia[§]
School of Economic Sciences
Washington State University
Pullman, WA 99164

December 1, 2020

Abstract

This paper examines the role of firms' uncertainty about consumers' environmental concerns in the emergence of greenwashing. We consider a signaling game where a firm (either green or brown) chooses whether to acquire a green label to signal its type to consumers (either green or brown). Uncertainty stems from two sources: consumers are uninformed about the firm's type, and firms are uninformed about the consumer's type. We examine under which conditions information transmission arises in equilibrium, showing that it critically depends on: (i) the proportion of green consumers; (ii) the valuation premium that green consumers assign to the green good, relative to brown consumers; (iii) the labeling cost for the brown firm; and (iv) the penalty that firms suffer after practicing greenwashing. We also identify pooling equilibria in which greenwashing is promoted. Finally, we find that bilateral uncertainty, relative to the case in which only the consumer is uninformed (unilateral uncertainty), may hinder greenwashing.

Keywords: Signaling game; bilateral uncertainty; green label; greenwashing.
JEL classification: D81, D82, L15, Q50.

[*]We would like to especially thank Carmen Arguedas, Mads Greaker, Jenny De Freitas, and Santiago Rubio for their insightful comments and suggestions at the 3rd AERNA Workshop on Game Theory and The Environment in Valencia, Spain. We also are greatful to Charles F. Mason and Yi Li for their useful feedback and to participants at the 94th WEAI Annual Conference in San Francisco, CA, and at the 15th WEAI International Conference in Tokyo, Japan, for comments and suggestions.
[†]118 Karp Hall, Union College, Schenectady, NY 12308. E-mail: garridom@union.edu.
[‡]101B Hulbert Hall, WSU, Pullman, WA 99164. E-mail: anaespinola@wsu.edu.
[§]103H Hulbert Hall, WSU, Pullman, WA 99164. E-mail: fmunoz@wsu.edu.

1 Introduction

Consumers are generally becoming more environment-conscious in their purchasing decisions. For instance, almost half of US consumers search for environmental information when deciding what good to buy, as reported in Cone Communications (2013). US sales of food and beverages, household and personal care products with sustainable attributes grew four times faster than sales of conventional products from 2014 to 2017; and sales of sustainable products are expected to make up 25% of total store sales by 2021 (Nielsen, 2018). Environmental characteristics of products are, however, not publicly observable to consumers. Green firms —which, for generality, we understand as those using environment-friendly inputs or production processes— usually make green claims in the product's label to overcome this information asymmetry, explaining why the number of firms using green labels has considerably increased in the last fifteen years.[1]

This information asymmetry and firms' incentives to increase sales constitute a breeding ground for greenwashing in the form of false or misleading green labels; as reported in Delmas and Burbano (2011). Terrachoice (2010) found that 32% of a total of 5,296 home and family green products in the US and Canada used false labels. Greenwashing may, however, be penalized by consumers, environmental activists, and non-governmental organizations (NGOs), reducing firms' incentives to include false information about their product's characteristics; see Lyon and Maxwell (2011), Lyon and Montgomery (2015), Berrone et al. (2017), and Garrido et al. (2020).

While these models predict firm's behavior, they assume that firms perfectly observe consumer's concerns for environmental features in their product. In other words, only consumers are uninformed about the firm's type (green or brown) while firms are perfectly informed (unilateral uncertainty). In most settings, however, a firm can only estimate the consumer's environmental concerns, but these estimates do not fully resolve firm's uncertainty. For instance, a survey conducted by Nielsen (2018) shows that millennials assign a higher value than Baby Boomers to products that make social responsibility claims (80% vs. 48%). Therefore, even when firms face a millennial consumer, they still have uncertainty about her environmental preferences (e.g., 20% of surveyed millenials do not value products with social responsibility claims).

We study a setting where firms choose whether or not to label its product in a context of bilateral uncertainty, since both firm and consumer are uninformed about each other's type. The consumer's environmental concerns are critical for the firm when deciding whether to use a green label, since they affect her decision to purchase the good. In addition, when the firm practices greenwashing, it suffers a reputation penalty. For example, Volkswagen's greenwashing scandal in 2015 significantly impacted its profits and reputation. Bachmann et al. (2019) report that VW's end-of-day stock price fell by 33% in the two trading days following the scandal, still being 24% lower than pre-scandal closing prices by the end of 2016, and that social media discussions regarding VW shifted after the scandal, with positive sentiments declining and negative sentiment increasing.

[1] Over 450 types of green labels are currently used in the world (almost half of them in the US) under different schemes (public vs. private, mandatory vs. voluntary) and degree of sustainability or stringency; see Gruere (2013) and Ecolabel Index (2019).

[2] Furthermore, this reputation penalty may depend on the consumer's type. Indeed, consumers with strong environmental concerns tend to be more active at NGOs and online platforms, reducing the firm's future sales, relative to consumers with low environmental concerns.

Our paper examines how this additional layer of uncertainty between firms and consumers affects information transmission and the emergence of greenwashing. In particular, we consider a signaling game where a firm (either green or brown) decides whether to use a green label without observing the consumer's environmental concern. Consumers assign a valuation premium to the green good, but those with strong environmental concerns (which we refer as green consumers) assign a larger premium than consumers with weak environmental concerns (brown consumers). Consumer's types can then be asymmetric, if the green consumer assigns a strictly larger premium to the green good than the brown consumer does, or symmetric, if both types assign the same premium. When they are asymmetric, the above assumptions about consumer's valuations generate different price regions depending on whether the brown consumer purchases any type of good (when prices are relatively low), does not buy it (when prices are relatively high), or purchases it only after inferring that the good is green (when prices are intermediate).

As a benchmark, we first show that, under complete information, firms do not need to rely on labels to convey their type to consumers. However, under incomplete information (bilateral uncertainty), we demonstrate that a separating equilibrium can be sustained where the green firm labels its product to signal its type to the green and brown consumer. This equilibrium exists, in particular, when prices are relatively low and the proportion of green consumers is sufficiently high, thus inducing the green firm to label its product, but relatively low to prevent the brown firm from mimicking this labeling strategy. A similar argument applies when prices are relatively high. When prices are intermediate, nonetheless, a high proportion of green consumers is required since it increases the expected penalty that the brown firm suffers from greenwashing, preventing this type of behavior.

Our results also identify how the separating equilibrium is affected by parameter changes. If the proportion of green consumers increases, this equilibrium, and thus information transmission, holds under larger conditions when the green consumer penalizes greenwashing more significantly than the brown consumer does, that is, when the penalty differential is large enough. Otherwise, the separating equilibrium can only be sustained under more restrictive conditions. Our findings can be evaluated at settings of unilateral uncertainty, where the firm is certain of facing green consumers, where we show that a separating equilibrium arises only when the penalty differential is large enough, but cannot arise otherwise. In other words, improving the firm's information about the consumer's type shrinks the range of parameter values where information transmission emerges in equilibrium. In contrast, when the firm faces brown consumers with certainty, a separating equilibrium cannot be supported since the brown firm has strong incentives to label given that

[2] Similarly, the Alva Group (2016) documents that positive recommendations of VW dropped by an average of 67%. In April 2016, VW recorded its first annual loss in more than 20 years for 2015 (The Guardian, 2016) and by the end of the year 2015, VW's reputation dropped by 30% in Germany and other European countries and by 12.5% in the US (Reputation Institute, 2016).

greenwashing is less penalized by brown consumers. Intuitively, complete certainty about the consumer's type prevents labeling from serving as an informative tool in this context.

The separating equilibrium can, however, arise under larger conditions if the brown firm faces a higher cost of labeling or if the penalty differential from greenwashing increases. Intuitively, when the brown firm faces a higher "overall cost" from greenwashing, either in terms of its labeling cost or penalty, it becomes less attracted to label. A similar argument applies when consumers become more asymmetric in their premium to the green good, since that makes the brown consumer less responsive to labels, thus shrinking the parameter values where the separating equilibrium can be sustained.

We then identify a pooling equilibrium where both types of firm label its product, thus concealing information from uninformed consumers. This equilibrium is more likely to arise when the brown firm's overall cost of greenwashing is relatively low (in terms of labeling, penalty differential, or both) since this type of firm is more attracted to label. When the overall cost of greenwashing is high, the pooling equilibrium can only be sustained if the brown type is responsive to labels (when prices are intermediate) and when the proportion of green consumers is low enough (which decreases the firm's expected penalty).

We also develop comparative statics for the pooling equilibrium. First, we show that, for relatively high and low prices, a larger proportion of green consumers expands the set of parameter values supporting the pooling equilibrium, and thus greenwashing. However, when prices are intermediate, a larger proportion of green consumers may hinder greenwashing. Our results then suggest that societies that experience an increase in the proportion of green consumers may not necessarily hinder greenwashing and, instead, may observe firms practicing more greenwashing.[3]

A higher labeling cost and/or penalty differential shrinks, as expected, the range of parameters sustaining greenwashing. Finally, when the premium that the brown consumer assigns to the green good increases, approaching that of the green consumer, the brown firm becomes more attracted to greenwashing since the brown consumer is more responsive to labels, especially when prices are high. Therefore, when consumers become more symmetric in their premium, greenwashing is facilitated. In other words, policies that provide information about the environmental benefits of certain ingredients or production processes —often aimed at increasing the premium that consumers assign to green goods— may actually promote greenwashing if brown consumers increase their premium more significantly than green consumers do.

Our results suggest that regulations only requiring green firms to label their products can help information transmission, making the separating equilibrium more likely to arise. However, we show that the pooling equilibrium in which firms conceal their type from consumers can still be sustained, being unaffected by mandatory labeling. Nonetheless, when greenwashing fines are set at extremely high levels, only the separating equilibrium can be supported, although both types of firm earn lower profits than under complete information.

[3]For instance, if two countries exhibit similar penalties for greenwashing, we should observe this practice more frequently occurring in the country with a higher proportion of green consumers.

Our findings also provide different policy implications. When the proportion of green consumers is high, we predict that a separating equilibrium exists. In this context, regulators seeking to facilitate information transmission need to increase the labeling costs of brown firms or the penalties to companies practicing greenwashing only when prices are intermediate; otherwise, when prices are relatively low or high, there is no need of intervention. A similar argument applies when green consumers are unlikely, since the separating equilibrium cannot be sustained, implying that increasing labeling costs or penalties is ineffective at facilitating information transmission. In contrast, when the proportion of green consumers is intermediate (relatively heterogeneous populations), a policy increasing the above costs expands the conditions where the separating equilibrium is supported, thus facilitating information transmission.

These are common policy recommendations in the signaling literature assuming unilateral uncertainty (where firms are perfectly informed about the consumer's type) but may become unnecessary in a context of bilateral uncertainty; namely, when green consumers are relatively likely or unlikely and prices are not intermediate. These results are then particularly useful in markets in which penalties are difficult to implement, or in which government agencies cannot easily distinguish between firm's types, thus not being able to increase the labeling cost of the brown firm alone.

Related Literature. Our paper connects to the literature examining the use of green labels, or eco-labels, as signals to credibly convey information to environment-conscious consumers in a context of asymmetric information —see Yokessa and Marette (2019) for an extensive review of theoretical and empirical studies— and how this information setting may favor firms' practice of greenwashing in equilibrium; see Lyon and Montgomery (2015) for a recent analysis of this phenomenon. Previous studies mainly focus on the role of prices and signaling costs as mechanisms to facilitate information transmission. Mahenc (2008) finds that prices can act as signals to green consumers if the green product is more costly to produce than the brown. Volle (2017) analyzes the distortions created by price signaling and shows the emergence of an uninformative equilibrium promoting greenwashing when prices are the only signal under consideration.

Hamilton and Zilberman (2006) and Mahenc (2017) find that certification costs effectively reduce greenwashing if they are large enough to deter the brown firm from mimicking the green firm labeling strategy. Bottega and De Freitas (2019) assume imperfect certification, allowing the monitoring process to produce false positives (a low-quality product receiving a high-quality certificate) and false negatives (a high-quality product failing the certification test). They find that an informative equilibrium can be sustained if the rate of false negatives is sufficiently low. Mason (2011) also considers a model of imperfect certification where, as in our setting, green firms are more likely to pass the certification test.[4] He shows that increasing certification costs can reduce green profits and reduce welfare. None of these articles, however, consider how expected penalty costs from greenwashing promote information transmission, which arises in a pooling equilibrium, and how the presence of bilateral uncertainty between consumers and firms affects equilibrium results.

[4]However, his model allows the monitoring process to produce false positives (a brown product receiving the certificate) and false negatives (a green product failing the certification test).

The role of penalty costs to discourage false signaling is, however, less explored; as recognized by Connelly et al., (2011). Lyon and Maxwell (2011) introduce the role of penalties in the form of the threat from an activist. Arguedas and Blanco (2014) find that corporate social responsibility (CSR) greenwashing only arises with intermediate fines in markets with conventional and uncertified CSR products. In addition, they show that the magnitude of these penalties affects consumers' beliefs about the credibility of CSR claims; if penalties are low, consumers anticipate self-reported CSR practices as fraudulent. More recently, Garrido et al. (2020) show that significant greenwashing penalties help certifications to disseminate information about a product where consumers do not observe product features but firms observe the type of consumer they face. Our paper considers a more general setting where firms and consumers cannot observe each other's types, helping us understand if an additional layer of uncertainty facilitates or hinders greenwashing practices.

Several papers analyze labeling under unilateral uncertainty, where firms can perfectly observe consumer's types. Ibanez and Grolleau (2008) assume that firms can invest in technology (green or brown) during the first stage of the game. In the second stage, firms choose whether to label their products; and, in the third stage, firms set their prices assuming that consumers updated their beliefs after observing the firm's labeling strategy. Daughety and Reinganum (2008) consider a model where firms can disclose the quality of their products (earning complete information profits) at a fixed cost or, instead, use their prices to signal their product quality. They examine under which parameter conditions firms choose to disclose their product quality in equilibrium and whether this disclosing strategy is socially optimal. Harbaugh et al. (2011) allow for firms to acquire a certificate from a third-party agency to signal their quality if it exceeds the agency's standard. Consumers are, however, uninformed about both the firm's quality and the agency's standard.[5] Heyes et al. (2020) also study a setting where firms observe consumer types. In their context, firms can signal their type using two labels (one more stringent in environmental terms than the other) but consumers only know which label is the most stringent if they choose to pay a learning cost, examining how results are affected if this cost is subsidized.

Papers considering bilateral (or two-sided) uncertainty between firms and consumers mainly focus on bargaining situations —see Chatterjee and Samuelson (1983) and Cramton (1992)— and on adverse selection scenarios, Gale (2001). However, Baksi and Bose (2007) consider a model of bilateral uncertainty with green and brown firms and with a distribution of consumers, all assigning a larger value to the green than the brown good. The premium that each consumer assigns to the green good is uniformly distributed in $[0, 1]$ implying that, as in our model, firms do not observe the consumer's type before making their labeling decision. However, this uniform distribution does not allow for a direct comparison between the bilateral and unilateral uncertainty settings. In contrast, our model helps understand how more certain consumer's types affect the firm's labeling strategy and, ultimately, its ability to practice greenwashing.

[5] They study the "Groucho effect,"namely, consumers lower their beliefs about the agency's standard when the firm acquires a certificate, as it means that the standard must be low for a given firm quality; and the "reverse Groucho effect," namely, that consumers update their beliefs about the agency's standard upward when the firm doesn't acquire the certificate, as it must indicate that the standard was high.

The next section describes the model and equilibrium results under complete information as our benchmark. Section 3 presents the time structure of the incomplete information game and identifies the separating and pooling equilibria. Section 4 provides a welfare analysis of our results, and section 5 analyses how labeling regulation might alter equilibrium results. Finally, section 6 provides a discussion of our results.

2 Model

Consider a signaling game between a firm and a consumer. The firm can be either green (G), if it has implemented an environmentally friendly production process, or brown (B) if it uses a dirty technology. Production costs are type-dependent, $C_i > 0$, where $i \in \{G, B\}$ denotes the firm's type, which allows for $C_G \geq C_B$ or $C_G < C_B$. The firm can follow two labeling strategies: to acquire a green label (L) or not (NL).[6] The cost of a green label is type-dependent, L_i, where $i \in \{G, B\}$ and $L_B \geq L_G \geq NL = 0$. Therefore, the cost of labeling can coincide for the brown and the green firm in the case of non-certified or self-reported green labels, $L_B = L_G$, or can be strictly higher for the brown firm in the case of a third-party certified label, $L_B > L_G$.

The brown firm faces an extra cost, k^j, in the form of an expected penalty (e.g., reputation damage) from greenwashing when using a green label, where $j \in \{G, B\}$ indicates the consumer's type, either green or brown, as discussed below. In particular, $k^G \geq k^B \geq 0$, indicating that the greenwashing penalty is weakly larger when facing a green than a brown consumer, e.g., larger loss of reputation. When firms do not acquire a green label, greenwashing does not occur, and they are not penalized by consumers.

To focus on the signaling role of labels, we consider that firms take price $p \geq 0$ as given.[7] When the green firm acquires a green label, its per-unit payoff is $p - C_G - L_G$ if consumer j buys, while the brown firm receives $p - C_B - L_B - k^j$. When firm i acquires a green label and the consumer does not buy, its payoff becomes $-C_i - L_i$. If the firm does not label, its payoff is $p - C_i$ if the consumer buys, and $-C_i$ if she does not. In addition, $p > L_i$, so both types of firm have incentives to acquire a green label.

The consumer can be either green (G) or brown (B) according to the valuation premium she assigns to the green good, as we describe next. Let V_i^j denote consumer j's valuation of the good, where $i \in \{G, B\}$ indicates the firm's type. If the consumer buys the product, her payoff is $V_i^j - p$, and zero otherwise. We assume that $V_G^j \geq V_B^j$ for every consumer j, that is, the valuation of a green good is higher for both green and brown consumers, which allows for the special case in which the consumer assigns the same value to both types of goods, $V_G^j = V_B^j$. However, the green consumer exhibits a higher valuation for the green good than the brown consumer does (i.e., $V_G^G > V_G^B$).

The above assumptions on consumer's valuations yield four price regions, as shown in figure 1.

[6] The acquisition of a green label is assumed not to be mandatory. Otherwise, labels could not act as a signal. We discuss this setting in Section 5.

[7] This may happen in highly competitive industries and in regulated industries where companies are required by law to charge a specific price p (such as utilities).

For simplicity, V_B^G is normalized to zero. If the price of the good is relatively low (i.e., $p \leq V_B^B$ in Region A), and the green consumer faces a green (brown) good, she purchases (does not purchase) since $V_G^G > p$ ($V_B^G = 0 < p$, respectively). In contrast, the brown consumer purchases both the brown and green good since they are sufficiently inexpensive, i.e., $V_B^B > p$ and $V_G^B > p$, respectively. In Region B, both types of consumer only buy the green good. However, in Region C, the green consumer only buys the green good, while the brown consumer does not buy any type of good. Finally, if the price is sufficiently high (Region D), no type of consumer buys the green or brown good. Since the consumer does not purchase the product in this region regardless of her type and beliefs, our subsequent analysis focuses on Regions A-C.

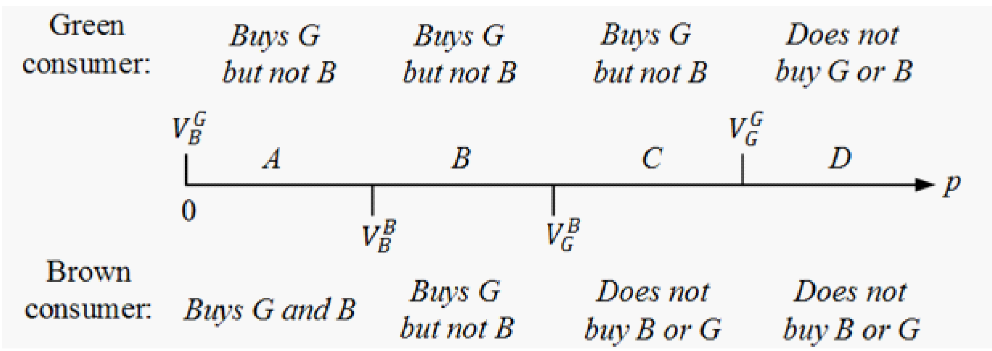

Figure 1. Price regions.

Therefore, the difference $V_G^j - V_B^j$ measures the valuation "premium" that consumer j assigns to the green relative to the brown good. The premium ratio

$$PR = \frac{V_G^B - V_B^B}{V_G^G - V_B^G}$$

then helps us compare the premium of the brown consumer relative to the green consumer's. For instance, when $PR = 0$, the brown consumer does not assign any premium to the green good (that is, $V_G^B = V_B^B$), indicating that both consumer's types are rather asymmetric in their premium for the green good. In contrast, when $PR = 1$, brown and green consumers assign the same premium to the green good.

The next lemma presents, as a benchmark, equilibrium behavior in a complete information setting where both consumer and firm can perfectly observe the other player's type. In particular, the firm chooses whether or not to acquire a green label. The consumer, observing the firm's labeling strategy and its type, responds buying or not buying the product. Proofs are relegated to the Appendix.

Lemma 1 (Complete information). *Under complete information, both types of firm do not label. Consumer j buys from firm i if her valuation for this good, V_i^j, satisfies $V_i^j \geq p$.*

Under complete information firm's type is known. Thus, no firm needs to spend resources on acquiring a green label, since purchasing decisions are based solely on the price and valuation of the good but are unaffected by labeling. In the following sections we show how the presence of uncertainty between the firm and the consumer leads one or both types of company to acquire a green label to induce purchases.

3 Equilibrium analysis

The time structure of the incomplete information game is the following, which assumes common knowledge among all players:

1. Nature selects the firm's type, either green with probability q, or brown with probability $1-q$. Nature also selects the consumer's type, either green with probability β, or brown with probability $1-\beta$.

2. The firm privately observes its type, but does not observe the consumer's type, and decides whether or not to acquire a green label.

3. The consumer also privately observes her type, but does not observe the firm's type. Upon observing the firm using a green label (no label), she updates her prior belief q to $\mu \equiv prob(G|L)$ ($\gamma \equiv prob(G|NL)$, respectively), where $\mu, \gamma \in [0,1]$. The consumer then responds buying or not buying the good.

We examine conditions under which separating and pooling Perfect Bayesian Equilibria (PBE) arise. All equilibria presented survive Cho and Kreps' (1987) Intuitive Criterion. For compactness, we normalize the penalty from brown consumers so that $K^B \equiv k^B - k^B = 0$ and $K^G = k^G - k^B \geq 0$. We can then interpret $K^G > 0$ as the penalty differential from green consumers, relative to brown consumers, while $K^G = 0$ indicates a zero-penalty differential, meaning that the penalty from both consumer's types coincides.

3.1 Separating equilibrium

Proposition 1. *A separating PBE in which the green firm acquires a label and the brown firm does not label can be sustained if and only if the following conditions hold in each price region:*

1. *The green and brown consumer buy after observing a green label:*

   (a) *Region A: $\beta_1(p) \leq \beta \leq \beta_2(p)$, where $\beta_1(p) \equiv \frac{L_G}{p} > 0$ and $\beta_2(p) \equiv \frac{L_B}{p-K^G} > 0$. In addition, cutoff $\beta_1(p)$ satisfies $0 < \beta_1(p) < 1$ under all admissible parameter conditions, while cutoff $\beta_2(p)$ satisfies $\beta_2(p) \leq 1$ if $p \geq L_B + K^G$.*

   (b) *Region B: $\beta \geq \beta_3(p)$, where $\beta_3(p) \equiv \frac{p-L_B}{K^G} > 0$. Furthermore, cutoff $\beta_3(p)$ satisfies $\beta_3(p) \leq 1$ if $p \leq L_B + K^G$.*

2. *Region C: The green consumer buys after observing a green label while the brown consumer does not if* $\beta_1(p) \leq \beta \leq \beta_2(p)$.

Figure 2 summarizes the equilibrium results in Proposition 1, where the 45°-line represents $L_B + K^G = p$. Parameter values above this line capture relatively high expected penalties, $L_B + K^G > p$, while values below the line indicate the opposite. When prices are sufficiently low (Region A where $p \leq V_B^B$, as presented in case 1a), the separating equilibrium (SE) can be sustained when the proportion of green consumers, β, is sufficiently high to induce the green firm to label its product (since green consumers only buy after observing a label), $\beta \geq \beta_1(p)$; but relatively low to prevent the brown firm from mimicking this labeling strategy, $\beta \leq \beta_2(p)$. Hence, a heterogeneous population (composed by green and brown consumers) can promote information transmission if reputation damage is low. Above the main diagonal, the condition on the proportion of green consumers, β, becomes less demanding. Intuitively, the penalty is high enough, leading the brown firm to not label its product regardless of the proportion of green consumers.

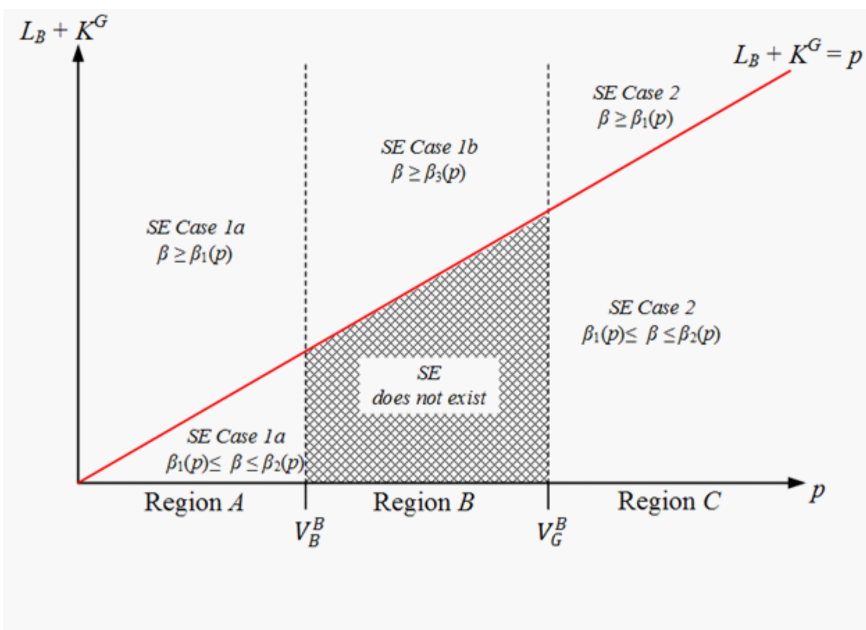

Figure 2. Separating PBE $L^G N L^B$.

In Region B (case 1b), the brown firm is more attracted to mimic the green firm's labeling strategy since prices are higher than in Region A, and both types of consumers buy the good only after observing a label. In this setting, the SE cannot be sustained if penalties are relatively low (below the 45^0-line). However, when penalties are high, the brown firm does not label if the proportion of green consumers is sufficiently high, $\beta \geq \beta_3(p)$, since the penalty becomes more likely.[8]

When prices further increase in Region C, brown consumers do not purchase any type of good. When the brown firm faces a large proportion of green consumers, $\beta \geq \beta_1(p)$, it does not have incentives to greenwash if the reputation damage is sufficiently high, $p < L_B + K^G$. Therefore, information transmission arises under the same conditions as in Region A. Nonetheless, cutoff $\beta_1(p)$ decreases in p, expanding the set of β that support the SE, above the 45^0-line. However, below this line the SE shrinks since the difference $\beta_2(p) - \beta_1(p)$ decreases in p. Intuitively, in this region, only green consumers buy the product after observing a label, giving the brown firm incentives to mimic.

We next show that the SE with the opposite labeling strategies cannot be sustained.

Proposition 2. *A separating PBE in which the green firm does not label and the brown firm acquires a label cannot be sustained in any price region.*

In this case, an informative equilibrium in which the brown firm acquires a label cannot be supported given that this firm would be revealing its type. In this situation, an unlabeled good is interpreted as being produced by a green firm. Thus, a label induces: (i) green consumers to not buy the good at any price since $V_B^G = 0 < p$; (ii) brown consumers to buy the labeled good if the price is low (i.e., $V_B^B \leq p$, in Region A); and (iii) brown consumers to not buy otherwise (Regions B and C). Therefore, a brown firm has incentives to not label its good.

3.1.1 Separating equilibrium - Comparative statics

Proportion of green consumers. Let us first examine the effect of an increase in the proportion of green consumers. In Region A, the existence of the SE only depends on cutoffs $\beta_1(p)$ and $\beta_2(p)$. When $\beta < \beta_1(p)$, this equilibrium cannot be sustained, since a low proportion of green consumers implies that the brown firm's expected penalty from greenwashing is too low. When $\beta_1(p) \leq \beta \leq \beta_2(p)$, the SE can be supported for all values of K^G; as graphically represented in points above and below the 45°-line in figure 2. Finally, when green consumers are sufficiently prevalent, $\beta > \beta_2(p)$, the SE exists only when the greenwashing penalty is high enough (parameter values above the 45°-line). A similar argument applies to Region C since it is sustained under the same conditions for β. Finally, in Region B, an increase in β makes the SE more likely to arise.

Cost of labeling. Our findings help predict how greenwashing practices are affected when labeling becomes more costly for the brown firm. This may happen when labeling must be done through a third-party certifying agency that requires a long and detailed process, making it almost impossible for the brown firm to receive a label (that is, $L_B \to +\infty$). Inspecting the upper area of figure 2, we observe that the SE can be sustained under less demanding conditions for β. Specifically, we only need $\beta \geq \beta_1(p)$ since all other cutoffs of β in Proposition 1 are not binding. Intuitively,

[8] Condition $\beta \geq \beta_3(p)$ is binding if the penalty is sufficiently high, $p < L_B + K^G$, as depicted in points above the 45°-line. In contrast, when $p \geq L_B + K^G$, cutoff $\beta_3(p)$ satisfies $\beta_3(p) > 1$, implying that condition $\beta \geq \beta_3(p)$ cannot hold.

labeling is unaffordable for the brown firm and, hence, the SE exists as long as the green firm finds it profitable to label. In contrast, when $L_B = L_G = 0$, cutoffs $\beta_1(p) = \beta_2(p) = 0$ and $\beta_3(p) = \frac{p}{K^G}$, thus shrinking the region of parameter values sustaining the SE.

Penalties from consumers. A similar argument applies when the penalty differential, K^G, increases. In this context, firms are more significantly penalized by green than brown consumers, making greenwashing less attractive, and expanding the set of parameter values sustaining the SE. However, when K^G decreases (or becomes zero), green and brown consumers similarly penalize greenwashing, making it more attractive to label, ultimately shrinking the set of parameter values for which a SE exists.[9] Overall, an increase in either L_B or K^G facilitates information transmission and hinders greenwashing.

Changes in consumer tastes. Finally, we can evaluate how our equilibrium results are affected by a change in consumer's valuations using the premium ratio, $PR = \frac{V_G^B - V_B^B}{V_G^G - V_B^G}$. When $PR = 0$, brown and green consumer are extremely asymmetric in their premium. In this case, Region B cannot be sustained, and only equilibrium findings in Regions A and C are supported, promoting information transmission.[10] In contrast, when $PR = 1$, consumer's types are symmetric in their premiums, even if their valuations of specific goods can differ. In this context, Region B spans all figure 1, preventing Regions A and C from occurring. Intuitively, the brown consumer is willing to pay a similar premium for the green good than the green consumer does, which attracts the brown firm to greenwash, and ultimately hinders information transmission.

3.2 Pooling equilibrium

We next analyze which pooling equilibria (PE) can be sustained. Define probability cutoff $\overline{q}^j(p) \equiv \frac{p - V_B^j}{V_G^j - V_B^j}$ for consumer $j \in \{G, B\}$. For simplicity, we restrict our attention to cases in which green firms are prevalent, that is, $q \geq \min\{\overline{q}^B(p), \overline{q}^G(p)\}$. In addition, when price lies in Region B we examine two cases: (i) area B-I, where $p \leq \widehat{p}$; and (ii) area B-II, where $p > \widehat{p}$ and cutoff $\widehat{p} \equiv \frac{V_B^B V_G^G}{V_G^G - V_G^B + V_B^B}$. When $p \leq \widehat{p}$, probability cutoffs satisfy $\overline{q}^B(p) \leq \overline{q}^G(p)$, which implies that, for a given prior belief q, green consumers are less likely to purchase the good in a pooling strategy profile where both types of firm use the same labeling strategy. The opposite argument applies if $p > \widehat{p}$ where probability cutoffs satisfy $\overline{q}^B(p) > \overline{q}^G(p)$.

Proposition 3. *A pooling PBE in which both types of firm acquire a label can be sustained if and only if the following conditions hold in each price region:*

1. *Region A: $\beta \geq \beta_2(p)$, equilibrium beliefs satisfy $q \geq \overline{q}^G(p)$ and off-the-equilibrium beliefs are $\gamma < \overline{q}^G(p)$.*

[9] Specifically, when $K^G = 0$, cutoff $\beta_2(p)$ decreases while $\beta_1(p)$ remains unaffected. In this context, Regions A and C sustain the SE under more restrictive conditions, but Region B cannot support the SE under any parameter values.

[10] In addition, when the green consumer assigns the same valuation to both types of good, $V_G^G = V_B^G$, Regions A-C collapse, leaving us with only Region D in figure 1 where no consumer purchases the good regardless of her beliefs.

2. *Region B-I:*

   (a) $\beta \geq \beta_2(p)$, *equilibrium beliefs satisfy* $q \geq \overline{q}^G(p)$ *and off-the-equilibrium beliefs are* $\overline{q}^B(p) \leq \gamma < \overline{q}^G(p)$.

   (b) $\beta \leq \beta_3(p)$, *equilibrium beliefs satisfy* $q \geq \overline{q}^G(p)$ *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^B(p)$.

   (c) $\beta \leq \beta_5(p)$, *equilibrium beliefs satisfy* $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^B(p)$, *where* $\beta_5(p) \equiv \frac{p-L_B}{p}$.

3. *Region B-II:*

   (a) $\beta \leq \beta_6(p)$, *equilibrium beliefs satisfy* $q \geq \overline{q}^B(p)$ *and off-the-equilibrium beliefs are* $\overline{q}^G(p) \leq \gamma < \overline{q}^B(p)$, *where* $\beta_6(p) \equiv \frac{p-L_B}{p+K^G}$.

   (b) $\beta \leq \beta_3(p)$,*equilibrium beliefs satisfy* $q \geq \overline{q}^B(p)$ *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^G(p)$.

   (c) $\beta \geq \beta_2(p)$, *equilibrium beliefs satisfy* $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^G(p)$.

4. *Region C:* $\beta \geq \beta_2(p)$, *equilibrium beliefs satisfy* $q \geq \overline{q}^G(p)$ *and off-the-equilibrium beliefs are* $\gamma < \overline{q}^G(p)$.

Figure 3 depicts the equilibrium results in Proposition 3. In Region A, prices are relatively low, inducing the brown firm to label only if the expected penalty is low (below the 45°-line) and the probability of facing a green consumer is high, $\beta \geq \beta_2(p)$, as this type of consumer purchases its good.[11]

In Region B-I, the brown firm labels if the penalty is sufficiently low, $p \geq L_B + K^G$, but green consumers are relatively likely, $\beta \geq \beta_2(p)$; as illustrated in case $2a$ of the figure. Intuitively, the expected penalty is low enough for this type of firm to mimic. When green consumers are unlikely, the brown firm mimics regardless of the penalty from greenwashing, K^G, as captured in cases $2b$ and $2c$.[12] In this case, purchase occurs if a label is observed. Hence, labeling induces the purchase of the (likely) brown consumers, and the brown firm does not incur penalties from greenwashing. Graphically, the PE is sustained both above and below the 45°-line.

Similarly, in Region B-II, the brown firm labels if the expected penalty is relatively low, namely when $p \geq L_B + K^G$ and $\beta \geq \beta_2(p)$; as depicted in case $3c$. However, since prices are more attractive in Region B-II than in B-I, penalty K^G can be higher and still induce greenwashing. When green

[11] Recall from Proposition 1 that cutoff $\beta_2(p)$ satisfies $\beta_2(p) \leq 1$ if and only if $p \geq L_B + K^G$. Otherwise, cutoff $\beta_2(p) > 1$, entailing that condition $\beta \geq \beta_2(p)$ of Proposition 3 (case 1) cannot hold above the 45^0-line.

[12] When K^G is sufficiently high, $K^G > p$, case $2b$ can be sustained under larger parameter conditions than case $2c$ since $\beta_5(p) > \beta_3(p)$.

consumers are unlikely, the brown firm labels regardless of the severity of the penalty K^G; as captured in cases $3a$ and $3b$, above and below the $45°$-line.

In Region C, prices are sufficiently high and the brown firm is willing to greenwash under larger penalties than in Region B-II. However, when penalties are sufficiently high (above the 45^0-line), the brown firm does not mimic since it now relies on the purchases of the green consumer alone, while in Region B both types of consumers buy the good. As a consequence, the PE cannot be sustained.

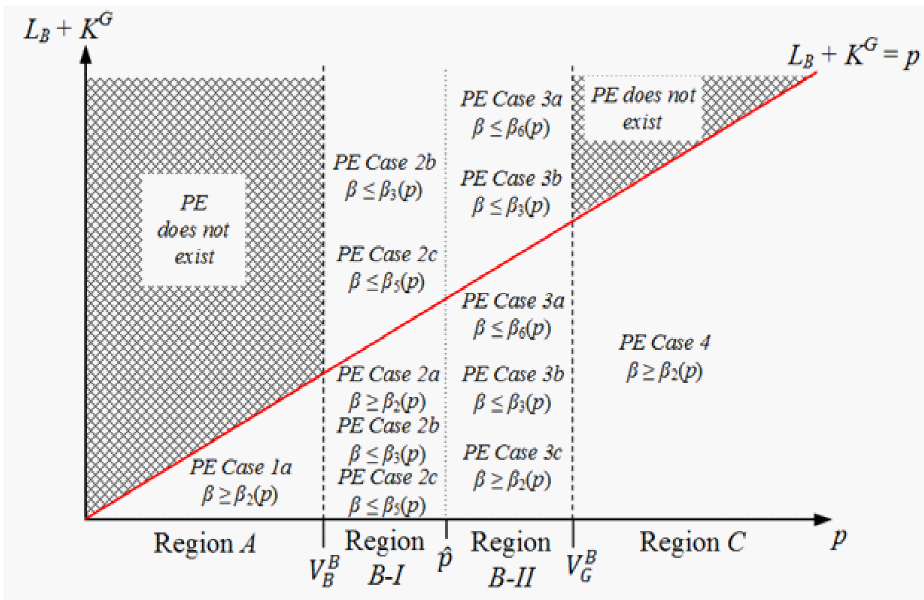

Figure 3. Pooling PBE $L^G L^B$.

We next discuss the PE in which neither type of firm acquires a green label.

Proposition 4. *A pooling PBE in which both types of firm do not label can be sustained if and only if the following conditions hold in each price region:*

1. *Region A: equilibrium beliefs satisfy $q \geq \overline{q}^G(p)$.*
2. *Region B-I:*
   (a) *equilibrium beliefs satisfy $q \geq \overline{q}^G(p)$,*
   (b) *$\beta \leq \beta_1(p)$, and equilibrium beliefs satisfy $\overline{q}^B(p) \leq q < \overline{q}^G(p)$.*
3. *Region B-II:*
   (a) *equilibrium beliefs satisfy $q \geq \overline{q}^B(p)$,*
   (b) *$\beta \geq \beta_4(p)$, where $\beta_4(p) \equiv \frac{p-L_G}{p}$, and equilibrium beliefs satisfy $\overline{q}^G(p) \leq q < \overline{q}^B(p)$.*

4. *Region C: equilibrium beliefs satisfy* $q \geq \overline{q}^G(p)$,

which hold for all off-the-equilibrium beliefs $\mu \in [0, 1]$.

When green firms are prevalent, $q \geq \overline{q}^G(p)$, and price lies in Region A (case 1), no type of firm has incentives to acquire a green label since both green and brown consumers are willing to buy the unlabeled good. In such a situation, firms make higher profit by saving the labeling cost, and the brown firm avoids the expected penalty K^G from greenwashing.

In Region B, this PE can be sustained when green firms are likely no matter the proportion of green consumers (cases 2a and 3a) under a similar argument than in case 1. In addition, this PE also arises in situations in which green firms are less prevalent and the proportion of green consumers is relatively low (case 2b, where $\beta \leq \beta_1(p)$) when prices are less attractive (Region B-I) and green consumers are less likely to purchase the good. In contrast, when prices are in Region B-II (in this case green consumers buy), the PE arises if green consumers are more prevalent, $\beta \geq \beta_4(p)$, as described in case 3b.

Finally, when prices are sufficiently high (Region C), brown consumers do not purchase the good regardless of their beliefs, leaving only green consumers as potential buyers. In this context, the above PE also emerges if green firms are likely, $q \geq \overline{q}^G(p)$. However, this condition is more demanding than in Region A since cutoff $\overline{q}^G(p)$ increases in price p. Therefore, this PE can emerge in contexts in which the green nature of firms is well known by consumers (i.e., strong private brand reputation) or the environmentally friendly production process is widespread.[13]

We next examine how the PE where greenwashing arises (Proposition 3) is affected by changes in the parameters.

3.2.1 Pooling equilibrium - Comparative statics

Proportion of green consumers. Let us first analyze how the proportion of green consumers affects the emergence of the PE in which both types of firm acquire a green label. In Region A, the existence of this PE depends on both the proportion of green consumers, β, and the greenwashing penalty, K^G. A sufficiently high penalty, $p < L_B + K^G$, deters greenwashing regardless of the proportion of green consumers. However, if this penalty is low (points below the 45°-line in figure 3), a minimum proportion of green consumers, $\beta \geq \beta_2(p)$, with strong beliefs ($q \geq \overline{q}^G(p)$), becomes a condition for this PE to arise. When this proportion decreases, however, the PE is less likely to emerge since green consumers are less prevalent (brown consumers buy in this region regardless of the signal).

A similar argument applies in Region C. However, cutoff $\overline{q}^G(p)$ is higher in this region, since it increases in price p, thus shrinking the parameter conditions under which this PE arises. In other words, a higher price requires the consumer to have stronger beliefs (i.e., higher q) about facing a

[13] In The Netherlands and Norway only 26% and 37% of consumers, respectively, look for environmental information relative to a 50% of consumers (on average) in other European countries (European Commission, 2014). This could be explained by consumers assuming that most firms use green technologies.

green firm to still purchase the good. Finally, in Region B, when the greenwashing penalty is low, an increase in β promotes greenwashing (see cases $2a$ and $3c$ in figure 3).

Cost of labeling. When the cost of labeling for the brown firm is sufficiently high, figure 3 indicates that the PE cannot be supported in any price regions. In particular, when $L_B \to +\infty$, the PE can only be sustained in Region B of figure 3, but conditions $\beta \leq \beta_3(p)$, $\beta \leq \beta_5(p)$, or $\beta \leq \beta_6(p)$ cannot hold since all these cutoffs become negative when L_B is sufficiently high. Therefore, when labeling is expensive for the brown firm, this type of company does not have incentives to mimic regardless of how attractive prices are, ultimately hindering the emergence of greenwashing under all parameter conditions.

Penalties from consumers. A similar argument applies when the penalty differential, K^G, increases. Firms are, in this setting, more penalized by green than brown consumers, making greenwashing less attractive, and shrinking the set of parameter values for which the PE can be sustained. In contrast, when K^G decreases, the PE can be supported under larger parameter values. For instance, when the penalties from green and brown consumers coincide, $K^G = 0$, all cases in Proposition 3 hold under less stringent conditions on the proportion of green consumers, β, except for case $2c$ which is unaffected.[14] In summary, the brown firm has less incentives to greenwash when either the labeling cost, L_B, the penalty differential from green consumers, K^G, or both, are sufficiently high.

Changes in consumer tastes. Finally, when the premium that the brown consumer assigns to the green good, $V_G^B - V_B^B$, increases, greenwashing becomes more attractive for the brown firm, and the uninformative PE holds under larger conditions. In the extreme case where the premium ratio is $PR = 1$, Regions A and C cannot be sustained. In contrast, when PR decreases, the brown consumer assigns a lower premium to the green good than the green consumer does (in the extreme $PR = 0$ when $V_G^B = V_B^B$), indicating that the brown firm becomes less attracted to greenwash.

Finally, we next examine how our SE and PE in Propositions 1 and 3, respectively, are affected when labeling costs or greenwashing penalties coincide.

Corollary 1. *When $L_B = L_G$ or $K^G = 0$, the SE (PE) is supported under more (less) restrictive parameter conditions.*

Intuitively, when both firms face the same labeling cost or the same greenwashing penalty from both consumer types, the green firm's separating effort becomes more costly, thus facilitating the emergence of pooling equilibria., ultimately hindering information transmission and promoting greenwashing.

[14] Specifically, in Region A, cutoff $\beta_2(p)$ is lower when $K^G = 0$ than when $K^G > 0$, expanding the parameter values satisfying $\beta \leq \beta_2(p)$ in this region. In Region *B-I*, cutoff $\beta_2(p)$ decreases, $\beta_3(p)$ increases, and $\beta_5(p)$ is unaffected, expanding the PE in cases $2a$ (which holds for a larger range of β) and $2b$ (which now holds for all values of β), while leaving case $2c$ unaffected. In Region *B-II*, cutoff $\beta_6(p)$ increases. Along with the above results, this means that cases $3a$ and $3c$ hold under a larger range of β, whereas case $3b$ holds for all values of β.

3.3 Comparison with unilateral uncertainty

Separating equilibrium. The following corollary evaluates our above results when the firm is certain about the type of consumer it faces.

Corollary 2. *When $\beta = 1$, the separating PBE can be sustained if and only if $p < L_B + K^G$ for all price regions. When $\beta = 0$, the separating PBE cannot be supported under any parameter values.*

In the extreme case where $\beta = 1$, the firm is certain about only facing green consumers and the SE can only be supported if the penalty from greenwashing is sufficiently high, $p < L_B + K^G$; graphically depicted in points above the 45°-line across all price regions. Therefore, information transmission arises under larger conditions when the firm is uncertain about consumer's types. Intuitively, the brown firm is less attracted to mimic the green firm's labeling strategy when it faces uncertainty, expanding the regions where the SE can be sustained. In contrast, when $\beta = 0$, the SE cannot be supported since the brown firm has incentives to label given that greenwashing is less penalized by brown than green consumers. In this context, complete certainty about consumer's type prevents labeling to serve as an informative tool.

Pooling equilibrium. We next evaluate our equilibrium results at the extreme cases where the firm is certain about the type of consumer it faces.

Corollary 3. *When $\beta = 1$, the pooling PBE of Proposition 3 can be sustained if and only if $p \geq L_B + K^G$ under all price regions. When $\beta = 0$, this pooling PBE can only be supported in Region B.*

Therefore, when the firm is certain of facing green (brown) consumers, greenwashing becomes more (less) attractive for the brown firm, expanding (shrinking, respectively) the parameter conditions where the PE can be sustained. However, when the PE arises independent of K^G (cases $2b$ and $2c$ in Region $B\text{-}I$, and cases $3a$ and $3b$ in Region $B\text{-}II$), a reduction in β to $\beta = 0$ helps promote greenwashing. Intuitively, when the brown firm is certain of facing brown consumers, it is not penalized from greenwashing. This is the only context in which a larger proportion of brown consumers promotes greenwashing.

3.4 Combining separating and pooling equilibria

We now seek to identify under which conditions on the proportion of green consumers, β, the SE (Proposition 1), the PE (Proposition 3), or both, can be sustained. This characterization depends on the ranking of cutoffs $\beta_1(p)$ to $\beta_6(p)$, as the following corollary describes.

Corollary 4. *If $p \geq L_B + K^G$, cutoffs are ranked as follows $0 < \beta_1(p) \leq \beta_2(p) \leq \beta_6(p) \leq \beta_5(p) \leq \beta_4(p) < 1 < \beta_3(p)$. If, instead, $p < L_B + K^G$, the cutoff ranking depends on the penalty differential, K^G, as follows:*

1. *If $p \geq K^G$, the cutoff ranking becomes* $0 < \beta_6(p) \leq \beta_5(p) \leq \beta_4(p) \leq \beta_3(p) < \beta_1(p) < 1 < \beta_2(p)$.
2. *If $2L_G \geq K^G > p$, the cutoff ranking becomes* $0 < \beta_6(p) < \beta_3(p) < \beta_5(p) \leq \beta_4(p) < \beta_1(p) < 1 < \beta_2(p)$.
3. *If $L_B + L_G \geq K^G > \max\{p, 2L_G\}$, the cutoff ranking becomes* $0 < \beta_6(p) < \beta_3(p) < \beta_5(p) < \beta_1(p) \leq \beta_4(p) < 1 < \beta_2(p)$.
4. *If $K^G > \max\{p, L_B + L_G\}$, the cutoff ranking becomes* $0 < \beta_6(p) < \beta_3(p) < \beta_1(p) \leq \beta_5(p) \leq \beta_4(p) < 1 < \beta_2(p)$.

Intuitively, when the penalty differential that the brown firm suffers after greenwashing, K^G, increases (moving from case 1 to 4 in Corollary 4), the SE expands; that is, it can be sustained under less stringent conditions on the proportion of green consumers, β. In contrast, greenwashing arises under more restrictive conditions on β. Alternatively, when the penalty differential is severe, the PE can only be supported if the proportion of brown consumers, $1 - \beta$, is sufficiently high. Given this ranking of cutoffs, we can now describe which equilibria coexist and in which regions this occurs.

Corollary 5. *Separating and pooling PBEs in Propositions 1 and 3 can only coexist in the following regions:*

1. *Region B-I:*
   (a) $p < L_B + K^G$ *(above 45^0-line):*
      i. *The SE in Proposition 1.1b and the PE in Proposition 3.2c coexist if* $\beta \in [\beta_3(p), \beta_5(p)]$ *and* $p < K^G$.
      ii. *The PE in Proposition 3.2b and the PE in Proposition 3.2c coexist if* $\beta \leq \min\{\beta_3(p), \beta_5(p)\}$.
   (b) $p \geq L_B + K^G$ *(below 45^0-line): The PEs in Proposition 3.2a, 3.2b, and 3.2c coexist if* $\beta \in [\beta_3(p), \beta_5(p)]$.
2. *Region B-II:*
   (a) $p < L_B + K^G$ *(above 45^0-line): The PEs in Proposition 3.3a and 3.3b coexist if* $\beta \leq \beta_6(p)$.
   (b) $p \geq L_B + K^G$ *(below 45^0-line): The PEs in Proposition 3.3a, 3.3b, and 3.3c coexist if* $\beta \in [\beta_2(p), \beta_6(p)]$.

Therefore, SE and PE cannot coexist when prices are extremely low or high (Regions A and C). However, when prices are relatively low (Region B-I) and the penalty differential K^G is high, both

SE and PE may coexist if the proportion of green consumers, β, is not extremely high; see cases $1a(i)$-(ii) in Corollary 5. Graphically, this occurs above the 45^0-line, where the SE of case $1b$ holds when β is sufficiently high, and the PE of cases $2b$ and $2c$ can be sustained when β is intermediate. Below the 45^0-line, the SE cannot be supported, but the three PEs in case 2 of Proposition 3 can arise if the proportion of green consumers is intermediate. Otherwise, only one of them emerges. When prices increase to Region B-II, the PEs in Proposition 3 (case 3) can coexist. Specifically, when the proportion of green consumers is relatively low (high), two (three) PEs coexist.

4 Welfare analysis

We now use our results from Corollary 5, identifying in which regions different equilibria may coexist, to find which of these PBEs yields the highest social welfare. For presentation purposes, cases are considered following the same order as in Corollary 5. We assume that the social planner cannot observe firm's or consumer's types, and computes the expected welfare as the sum of the consumer utility from the good (when a purchase occurs in equilibrium) and the firm's profit from selling the product.

Proposition 5. *In the regions where more than one PBE can be sustained, welfare levels are ranked as follows:*

1. *Region B-I:*
   (a) $p < L_B + K^G$ *(above 45^0-line):*
      i. *The SE in Proposition 1.1b yields a higher welfare than the PE in Proposition 3.2c if and only if* $\frac{q}{1-q} \geq \frac{(1-\beta)V_B^B - L_B}{\beta V_G^G}$.
      ii. *The PE in Proposition 3.2b generates a higher welfare than the PE in Proposition 3.2c if and only if* $\frac{q}{1-q} \geq \frac{K^G}{V_G^G}$.
   (b) $p \geq L_B + K^G$ *(below 45^0-line): The PEs in Proposition 3.2a and 3.2b yield the same welfare, which exceeds that in Proposition 3.2c if and only if* $\frac{q}{1-q} \geq \frac{K^G}{V_G^G}$.
2. *Region B-II:*
   (a) $p < L_B + K^G$ *(above 45^0-line): The PEs in Proposition 3.3a and 3.3b generate the same welfare.*
   (b) $p \geq L_B + K^G$ *(below 45^0-line): The PEs in Proposition 3.3a and 3.3b yield the same welfare, which exceeds that in Proposition 3.3c under all parameter values.*

The SE and PE only coexist when penalty differentials are large (case $1ai$). In this context, Proposition 5 identifies that the SE is welfare superior if the proportion of green firms is relatively high, i.e., $\frac{q}{1-q} \geq \frac{(1-\beta)V_B^B - L_B}{\beta V_G^G}$. This condition holds under larger conditions when the cost of labeling for the brown firm, L_B, increases, and when the green consumer's expected valuation for the green good, V_G^G, also increases. A similar argument applies if the proportion of green consumers, β, increases. However, when the brown consumer assigns a high valuation to the brown good, V_B^B, the separating equilibrium is less likely to generate a higher social welfare.

Still in a setting with high penalties, two PEs can arise, as shown in case $1aii$ of Corollary 5. In this scenario, Proposition 5 finds that the PE where both types of consumer purchase the good (Proposition 3.2a) yields a larger welfare than that where only the brown consumer buys (Proposition 3.2b) if $\frac{q}{1-q} \geq \frac{K^G}{V_G^G}$. This condition is more likely to hold when the greenwashing penalty decreases and when the green consumer's valuation for the green good increases. A similar argument applies in a setting of low greenwashing penalties (case $2b$), where the PE inducing the purchase of both types of consumer leads to a higher social welfare.

4.1 Market inefficiencies and lemons

We next study the existence of socially inefficient purchases of the green good.

Separating equilibrium. Our above results indicate that, if a SE can be sustained, the green and brown consumer purchase the good under the same conditions as under complete information. Specifically, in Regions A and B the green and brown consumer purchase the good in the SE, thus coinciding with the buying profiles described in Lemma 1 and illustrated in figure 1. In Region C, only the green consumer buys the good in the SE, which also coincides with the buying profile of Lemma 1. Therefore, when parameter values sustain a SE, no lemons problem arises: sales that occur under complete information also happen under incomplete information. The green firm, however, earns a lower profit in the SE, since it needs to use labeling to signal its type. The brown firm, not labeling in the SE, earns the same profits as under complete information.

Pooling equilibrium. Let us now study market inefficiencies in the PE where both types of firm label its product (Proposition 3). Under complete information, the green consumer purchases the green good in all Regions A through C, as depicted in figure 1. In the PE, the green consumer also purchases the green good, except for intermediate prices (specifically, in Region B-Ic), where she does not buy the good, thus giving rise to inefficiencies due to incomplete information. A similar argument applies when we examine this consumer's purchasing of the brown good, although in the opposite price regions. Under complete information, she did not buy the brown good regardless of its price, but in the PE she purchases it in all price regions, giving rise to inefficiencies, except in Region B-Ic. In summary, the green consumer purchases the green good under more stringent settings than complete information, but buys the brown good under larger conditions.

Regarding the brown consumer, recall that under complete information she only purchases the green good in Regions A and B, but does not in Region C. In the PE, she purchases the green good under the same conditions, except in Region B-IIc where she does not buy the good, thus leading to inefficiencies. In the case of a brown good, this type of consumer purchases it under complete information only in Region A, which still holds under the PE. However, in Regions B-I and B-IIa, b, she buys the brown good, leading to socially excessive purchases relative to complete information. Finally, in Region C, the brown consumer does not buy the brown good regardless of the information setting.

Therefore, in the PE both types of consumer purchase the green good under more stringent settings than under complete information, but buy the brown good under less stringent contexts. Overall, this means that, in the PE we can expect socially insufficient purchases of the green good, relative to the purchasing decisions that emerge under complete information, and socially excessive purchases of the brown good.[15]

5 Labeling regulations

Non-governmental organizations and consumer groups often request policy changes that increase the penalties to firms practicing greenwashing and/or forcing green firms to label their products. We next examine whether these policies promote or hinder information transmission.

Penalties on greenwashing. Our above results also help predict how firms' labeling strategies can be affected by regulation that facilitates legal actions against firms that may have practiced greenwashing. The effect of this type of regulation is equivalent to an increase in the penalty from greenwashing by green consumers (e.g., loss of reputation), but occurs when the firm faces both green and brown consumers. As depicted in the upper portion of figure 2, an increase in penalties facilitates information transmission. When this regulation makes it prohibitively expensive for the brown firm to label its product (which is equivalent to making penalties from consumers extremely high, $k^G, k^B \to +\infty$), the SE can be sustained if $\beta > \beta_1(p)$ under all price regions. In contrast, the PE of figure 3 cannot hold when the brown firm's fine from greenwashing is sufficiently high, thus hindering greenwashing.

While severe penalties facilitate information transmission, our results highlight that the SE still gives rise to inefficiencies relative to complete information, since the green firm spends resources labeling its product, which the firm saves under a complete information context.

Forcing green firms to label. An alternative regulation could require green firms to label their products, setting an extremely high fine if this type of firm is caught without a label, while allowing brown firms to choose whether or not to label. In this setting, the green firm is no longer a strategic player, collapsing all the analysis to the brown firm's incentives to label. The SE where the green (brown) firm labels (does not label) can be now sustained under larger parameter conditions. In particular, the SE arises in Regions A and C if $p \geq L_B + K^G$ and the proportion of green consumers satisfies $\beta \leq \beta_2(p)$. Relative to Proposition 1, we no longer need a minimum proportion of green consumers, $\beta \geq \beta_1(p)$, for the green firm to label. When the penalty is high, $p < L_B + K^G$, this SE also expands, since it can be supported for all values of β. The PE is, however, sustained under the same conditions as in Proposition 3, thus being unaffected by mandatory labeling.

[15] Our discussion focuses on the PE where both types of firm label (Proposition 3) rather than that where both types of firm do not label (Proposition 4). It is straightforward to show, however, that the regions where excessive or insufficient purchases emerge in the PE of Proposition 4, relative to complete information coincide with those examined here for Proposition 3.

Another version of the above policy could still require the green firm to label its product, but demand that the brown firm does not label its product or otherwise pay an extremely severe fine. In this context, equilibrium results will, of course, coincide with those under the SE, although they would hold under all price regions and proportions of green consumers. While consumers can infer the green firm's type upon observing a label, and thus information transmission is promoted, this policy yields a lower welfare than under complete information, since the green firm is forced to label its product just to distinguish it from the brown firm.

6 Discussion

Proportion of green consumers. We can summarize our equilibrium results in Propositions 1 and 3 as a function of the proportion of green consumers, β. When the firm is almost certain of facing brown consumers (low values of β) and prices lie in Region B, only the PE can be sustained, and thus greenwashing occurs; which holds for all penalties, i.e., $p < L_B + K^G$ and $p \geq L_B + K^G$. In this case, both types of consumer buy the product only when inferring it is green; thus providing both types of firm with incentives to label.

When the proportion of green consumers increases enough, both SE and PE can coexist when the penalty differential, K^G, is low; whereas only the SE exists when K^G is relatively high. In addition, as K^G increases, the range of β's supporting the SE expands. Therefore, when $p \geq L_B + K^G$, a society with a larger proportion of green consumers does not necessarily eliminate the risk of greenwashing. However, when $p < L_B + K^G$, a society with more green consumers facilitates information transmission, reducing the risk of greenwashing.

More symmetric consumers. Our results also help us predict how equilibrium behavior is affected by consumers who assign similar premiums to the green good. When PR increases, both consumer's types exhibit a strong preference for the green good, inducing the brown firm to practice more greenwashing, and ultimately hindering information transmission. Informally, when all types of consumer become more concerned about the environmental properties of a product, greenwashing is actually promoted. The opposite result arises when PR decreases since, in this setting, brown consumers exhibit a substantially lower premium towards the green good than green consumers do —including the case in which brown consumers do not assign any premium to green goods. Therefore, the brown firm has less incentives to greenwash, and information transmission emerges under larger conditions.

Cost of labeling. Our results suggest that increasing the labeling cost for the brown firm, L_B, by, for instance, requiring a long bureaucratic process with several steps, helps prevent greenwashing. Specifically, this occurs in price regions where the brown consumer's purchasing decision is unaffected by labels (Regions A and C). In contrast, when the brown consumer buys only after observing a label (Region B), a small increase in labeling cost for the brown firm, L_B, does not necessarily hinder greenwashing. A sufficiently high labeling cost ($L_B \to +\infty$), however, prevents the PE. Interestingly, the PE is affected by L_B but unaffected by L_G, while the SE expands when

L_G is low and L_B is relatively high, i.e., when the labeling cost differential $L_B - L_G$ increases. Intuitively, this suggests that, to promote information transmission in markets with bilateral uncertainty, regulators informed about the firm's type should make the labeling and certification process as automatic (difficult) as possible for the green (brown, respectively) firm, rather than only focusing on altering the labeling cost of one type of firm.

Contested markets. Our findings can also be presented according to price regions A-C. As discussed in Section 2, extremely low (high) prices may arise when the firm operates in a contested (uncontested) market facing (not facing) entry threats; graphically depicted in Regions A (C, respectively). In these settings, our results suggest that, when greenwashing penalties are relatively low, we may observe true labeling, as part of the SE, when the proportion of green consumers is relatively low, but greenwashing otherwise. However, when penalties are high, greenwashing is completely prevented. Intuitively, the brown firm cannot alter the brown consumer's purchasing decisions by labeling its product in Regions A and C, and thus practices (avoids) greenwashing when its expected penalty is sufficiently low (high, respectively). Summarizing, low prices make greenwashing less attractive, while high prices could induce it. Hence, severe penalties need to be in place in uncontested markets to deter greenwashing.

In contrast, intermediate prices can emerge when the firm faces entry threats but from a relatively inefficient rival that suffers from a cost disadvantage, as illustrated in Region B. In this context, we show that information transmission can only be sustained when greenwashing penalties are relatively high, while the PE where the brown firm conceals its type from consumers arises for all greenwashing penalties.

Changes in aggregate demand. A similar argument applies when the aggregate demand for the green good is extremely high, making region C more likely to arise. Specifically, a higher price facilitates the emergence of the separating equilibrium where only the green firm acquires a label and only green consumers respond purchasing the good. Intuitively, the price is sufficiently high that the brown consumer does not buy the good irrespective of her beliefs about the product. Since the brown firm anticipates that only green consumers would purchase the good in this region, it refuses from greenwashing because its expected penalty would be relatively high. In contrast, the pooling equilibrium is less likely to arise in this context since greenwashing becomes less profitable.

Policy implications. Regulations increasing the penalties from greenwashing, or making greenwashing more likely to be prosecuted, facilitate the emergence of the SE and hinder that of the PE. As a consequence, information about the firm's type flows more easily to uninformed consumers. However, as discussed in section 5, these regulations yield a lower social welfare than under complete information since the green firm incurs a labeling cost that could save otherwise. A similar argument applies to policies that force the green firm to label its product.

Further research. Our theoretical predication could be empirically tested in industries where consumers cannot observe firms' production process (green or brown) and firms are uncertain about consumer types. In this context, labeling strategies could be explained by the firm's type (which researchers must observe after years of operation), price, penalties from greenwashing; among others. When prices are extremely low or high, our results predict the emergence of the SE, which empirically entails that the firm's type is significant at explaining labeling decisions. In contrast, when prices are intermediate, our findings suggest that PE arise, which implies that the firm's type is insignificant at explaining labeling strategies. Finally, our model could be expanded by allowing for a single continuous distribution on consumer taste for the environmental attribute of the good. In this setting, the consumer would purchase the good when her preference for this attribute is sufficiently strong, and the firm chooses to label anticipating this cutoff purchasing decision.

7 Appendix

7.1 *Proof of Lemma 1*

Under complete information, the green and the brown consumer can perfectly observe the firm's type. As a consequence, they buy the good if the value they assign to it is greater than the price at which it is sold, i.e., $V_i^j > p$, where $i \in \{G, B\}$ denotes the firm's type and $j \in \{G, B\}$ represents the consumer's type. In Region A, the green consumer buys the green good while the brown consumer buys the brown and green good. In region B, both types of consumer only buy the green good. In Region C, only the green consumer buys the green good. The firm anticipates these purchasing decisions and does not label since labeling does not affect consumer's beliefs about firm's type.

7.2 *Proof of Proposition 1*

Price in Region A**:** Upon observing a label, the consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the green and the brown consumer buy since $V_G^G > p$, and $V_G^B > p$, respectively. And after observing no label, the green consumer does not buy since $V_B^G = 0 < p$, and the brown consumer buys since $V_B^B \geq p$. Anticipating the consumer's response, the green firm acquires a label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(p - C_G)$, yielding $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm does not label if and only if $\beta(-C_B) + (1 - \beta)(p - C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)$, implying $\beta \leq \frac{L_B}{p - K^G} \equiv \beta_2(p)$. We check that $\beta_1(p) \in [0, 1]$; $\beta_2(p) \geq 0$, and $\beta_2(p) \leq 1$ if $p \geq L_B + K^G$. In addition, we know $\beta_1(p) \leq \beta_2(p)$ since $p(L_G - L_B) \leq L_G K^G$ is satisfied because by definition $L_B \geq L_G$. Hence, this separating equilibrium can be sustained as a PBE if and only if $\beta \in [\beta_1(p), \beta_2(p)]$, where $\beta_2(p) \leq 1$ if $p \geq L_B + K^G$; $\beta \geq \beta_1(p)$ otherwise.

Price in Region B**:** Upon observing a label, the consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the green and the brown consumer buy since $V_G^G > p$, and $V_G^B \geq p$, respectively. And after observing no label, the green and the brown consumer do not buy since $V_B^G = 0 < p$ and $V_B^B < p$, respectively. Anticipating the consumer's response, the green firm acquires a label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(-C_G)$, implying $p \geq L_G$, which is satisfied by definition. Similarly, the brown firm does not label if and only if $\beta(-C_B) + (1-\beta)(-C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)$, implying $\beta \geq \frac{p-L_B}{K^G} \equiv \beta_3(p)$. We check that $\beta_3(p) \geq 0$ and $\beta_3(p) \leq 1$ if $p \leq L_B + K^G$. Therefore, this separating equilibrium can be sustained as a PBE if and only if $\beta \geq \beta_3(p)$ and $p \leq L_B + K^G$.

Price in Region C**:** Upon observing a label, the consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\mu = 1$. No label conveys the opposite information, i.e., $\gamma = 0$. After observing a label, the green consumer buys since $V_G^G \geq p$, and the brown consumer does not buy since $V_G^B < p$. And after observing no label, the green and the brown consumer do not buy since $V_B^G = 0 < p$ and $V_B^B < p$, respectively. Anticipating the consumer's response, the green firm acquires a label if and only if $\beta(p - C_G - L_G) + (1-\beta)(-C_G - L_G) \geq \beta(-C_G) + (1-\beta)(-C_G)$, yielding $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm does not label if and only if $\beta(-C_B) + (1-\beta)(-C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(-C_B - L_B)$, implying $\beta \leq \frac{L_B}{p-K^G} \equiv \beta_2(p)$. We know $\beta_1(p) \in [0,1]$, $\beta_2(p) \geq 0$, and $\beta_2(p) \leq 1$ if $p \geq L_B + K^G$. In addition, we know $\beta_1(p) \leq \beta_2(p)$. Therefore, this separating equilibrium can be supported as a PBE if and only if and $\beta \in [\beta_1(p), \beta_2(p)]$, where $\beta_2(p) \leq 1$ if $p \geq L_B + K^G$; $\beta \geq \beta_1(p)$ otherwise.

7.3 *Proof of Proposition 2*

Price in Region A**:** Upon observing no label, the consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the green and the brown consumer buy since $V_G^G > p$, and $V_G^B > p$, respectively. And after observing a label, the green consumer does not buy since $V_B^G = 0 < p$, and the brown consumer buys since $V_B^B \geq p$. Anticipating the consumer's response, the green firm does not label if and only if $\beta(p-C_G)+(1-\beta)(p-C_G) \geq \beta(-C_G-L_G)+(1-\beta)(p-C_G-L_G)$, yielding $\beta \geq \frac{-L_G}{p}$, which is satisfied since $\beta \in [0,1]$ by definition. Similarly, the brown firm acquires a label if and only if $\beta(-C_B-L_B)+(1-\beta)(p-C_B-L_B) \geq \beta(p-C_B)+(1-\beta)(p-C_B)$, implying $\beta \leq \frac{-L_B}{p}$, which does not hold since $\beta \in [0,1]$; and the brown firm deviates towards no label. Therefore, this separating equilibrium cannot be supported as a PBE.

Price in Region B**:** Upon observing no label, the consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the green and the brown consumer buy since $V_G^G > p$, and $V_G^B \geq p$, respectively. And after observing a label, the green and the brown consumer do not buy since $V_B^G = 0 < p$ and $V_B^B < p$, respectively. Anticipating the consumer's response, the green firm does not label if and only if $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(-C_G - L_G) + (1-\beta)(-C_G - L_G)$, implying $p \geq -L_G$, which holds. Similarly, the brown firm acquires a label if and only if $\beta(-C_B - L_B) + (1-\beta)(-C_B - L_B) \geq \beta(p - C_B) + (1-\beta)(p - C_B)$, yielding $p \leq -L_B$, which is not satisfied; and the brown firm deviates towards no label. Hence, this separating equilibrium cannot be supported as a PBE.

Price in Region C**:** Upon observing no label, the consumer updates her beliefs by Bayes' rule, inferring that the firm must be green, i.e., $\gamma = 1$. A label conveys the opposite information, i.e., $\mu = 0$. After observing no label, the green consumer buys since $V_G^G \geq p$, and the brown consumer does not buy since $V_G^B < p$. And after observing a label, the green and the brown consumer do not buy since $V_B^G = 0 < p$ and $V_B^B < p$, respectively. Anticipating the consumer's response, the green firm does not label if and only if $\beta(p - C_G) + (1-\beta)(-C_G) \geq \beta(-C_G - L_G) + (1-\beta)(-C_G - L_G)$, implying $\beta \geq \frac{-L_G}{p}$, which is satisfied since $\beta \in [0,1]$ by definition. Similarly, the brown firm acquires a label if and only if $\beta(-C_B - L_B) + (1-\beta)(-C_B - L_B) \geq \beta(p - C_B) + (1-\beta)(-C_B)$, yielding $\beta \leq \frac{-L_B}{p}$, which does not hold; and the brown firm deviates towards no label. Hence, this separating equilibrium can not be supported as a PBE.

7.4 *Proof of Proposition 3*

Price in Region A: Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium; and $\gamma \in [0,1]$ off-the-equilibrium. After observing a label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ since $V_B^G = 0$ by definition; and the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$. After observing no label, since the off-the-equilibrium beliefs satisfy $\gamma \in [0,1]$, the green consumer buys if and only if $\gamma \geq \overline{q}^G(p)$; and the brown consumer buys if $\gamma \geq \overline{q}^B(p)$. Since $p - V_B^B < 0$ and $V_G^B - V_B^B > 0$, $\overline{q}^B(p) < 0$; thus $q, \gamma \geq \overline{q}^B(p) \in [0,1]$ is always satisfied, and the brown consumer buys regardless of the message she observes.

We analyze two different cases depending on the consumer's beliefs:

Case 1: Consumer's beliefs are $q, \gamma \geq \overline{q}^G(p)$, that is, the green and the brown consumer buy regardless of the message they observe. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1-\beta)(p - C_G)$, yielding $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 1 cannot be sustained as a PBE.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$ and $\gamma < \overline{q}^G(p)$, that is, the green consumer buys after observing a label, and she does not buy otherwise; the brown consumer buys regardless of the message she observes. Anticipating the consumer' response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(p - C_G)$, which is satisfied if $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \geq \beta(-C_B) + (1-\beta)(p - C_B)$, which holds if $\beta \geq \frac{L_B}{p - K^G} \equiv \beta_2(p)$. We know $\beta_1(p) \in [0,1]$, $\beta_2(p) \geq 0$, $\beta_2(p) \leq 1$ if $p \geq L_B + K^G$, and $\beta_1(p) \leq \beta_2(p)$. Therefore, this pooling equilibrium in case 2 can be supported as a PBE if and only if $\beta \geq \beta_2(p)$ and $p \geq L_B + K^G$.

Price in Region B: Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium; and $\gamma \in [0,1]$ off-the-equilibrium. After observing a label, the green consumer buys if and only if $qV_G^G+(1-q)V_B^G-p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ since $V_B^G = 0$ by definition; and the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p-V_B^B}{V_G^B-V_B^B} \equiv \overline{q}^B(p)$. After observing no label, since the off-the-equilibrium beliefs satisfy $\gamma \in [0,1]$, the green consumer buys if and only if $\gamma \geq \overline{q}^G(p)$; and the brown consumer buys if and only if $\gamma \geq \overline{q}^B(p)$.

We analyze 12 different cases depending on the consumer's beliefs:

Region *B-I*: $p \leq \widehat{p} \rightarrow \overline{q}^B(p) \leq \overline{q}^G(p)$:

Case 1: Consumer's beliefs are $q, \gamma \geq \overline{q}^G(p)$, that is, the green and the brown consumer buy regardless of the message they observe. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1-\beta)(p - C_G)$, yielding $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 1 cannot be sustained as a PBE.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$ and $\overline{q}^B(p) \leq \gamma < \overline{q}^G(p)$, that is, the green consumer buys after observing a label, and does not otherwise; and the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(p - C_G)$, which implies $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \geq \beta(-C_B) + (1-\beta)(p - C_B)$, which yields $\beta \geq \frac{L_B}{p-K^G} \equiv \beta_2(p)$. Hence, this pooling equilibrium in case 2 can be supported as a PBE if and only if $\beta \geq \beta_2(p)$ and $p \geq L_B + K^G$.

Case 3: Consumer's beliefs are $q \geq \overline{q}^G(p)$ and $\gamma < \overline{q}^B(p)$, that is, the green and the brown consumer buy after observing a label, and do not otherwise. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(-C_G)$, yielding $p \geq L_G$, which is satisfied by definition. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \geq \beta(-C_B) + (1-\beta)(-C_B)$, which implies $\beta \leq \frac{p-L_B}{K^G} \equiv \beta_3(p)$. Therefore, this pooling equilibrium in case 3 can be supported as a PBE if and only if $\beta \leq \beta_3(p)$.

Case 4: Consumer's beliefs are $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ and $\gamma \geq \overline{q}^G(p)$, that is, the green does not buy after observing a label, and does otherwise; and the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1-\beta)(p - C_G)$, yielding $\beta \leq \frac{-L_B}{p}$, which does not hold since $\beta \in [0,1]$ by definition; and the green firm has incentives to deviate towards no label. Hence, this pooling equilibrium in case 4 cannot be sustained as a PBE.

Case 5: Consumer's beliefs are $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ and $\overline{q}^B(p) \leq \gamma < \overline{q}^G(p)$, that is, the green consumer does not buy regardless of the message she observes, and the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(p - C_G)$, implying $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 5 cannot be supported as a PBE.

Case 6: Consumer's beliefs are $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ and $\gamma < \overline{q}^B(p)$, that is, the green consumer does not buy regardless of the message she observes, and the brown consumer buys after observing a label and does not otherwise. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(-C_G)$, which holds if $\beta \leq \frac{p-L_G}{p} \equiv \beta_4(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(-C_B - L_B) + (1-\beta)(p - C_B - L_B) \geq \beta(-C_B) + (1-\beta)(-C_B)$, which is satisfied if $\beta \leq \frac{p-L_B}{p} \equiv \beta_5(p)$. We know that $\beta_4(p)$, $\beta_5(p) \in [0,1]$ since $p \geq L_i$. In addition, $\beta_4(p) \geq \beta_5(p)$ since $L_B \geq L_G$ by definition. Hence, this pooling equilibrium in case 6 can be sustained as a PBE if and only if $\beta \leq \beta_5(p)$.

Region *B-II*: $p > \widehat{p} \rightarrow \overline{q}^B(p) > \overline{q}^G(p)$:

Case 7: Consumer's beliefs are $q, \gamma \geq \overline{q}^B(p)$, that is, the green and the brown consumer buy regardless of the message they observe. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1-\beta)(p - C_G)$, implying $L_G \leq 0$, which does not hold; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 7 cannot be supported as a PBE.

Case 8: Consumer's beliefs are $q \geq \overline{q}^B(p)$ and $\overline{q}^G(p) \leq \gamma < \overline{q}^B(p)$, that is, the green consumer buys regardless of the message she observes; and the brown consumer buys after observing a label, and does not otherwise. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(p - C_G) + (1-\beta)(-C_G)$, which yields $\beta \leq \frac{p-L_G}{p} \equiv \beta_4(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \geq \beta(p - C_B) + (1-\beta)(-C_B)$, which implies $\beta \leq \frac{p-L_B}{p+K^G} \equiv \beta_6(p)$. We check $\beta_6(p) \in [0,1]$. In addition, we know $\beta_4(p) \geq \beta_6(p)$ since $p - L_G \geq p - L_B$ and $p < p + K^G$. Hence, this pooling equilibrium in case 8 can be supported as a PBE if and only if $\beta \leq \beta_6(p)$.

Case 9: Consumer's beliefs are $q \geq \overline{q}^B(p)$ and $\gamma < \overline{q}^G(p)$, that is, the green and the brown consumer buy after observing a label, and do not otherwise. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G) + (1-\beta)(-C_G)$, yielding $p \geq L_G$, which is satisfied by definition. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \geq \beta(-C_B) + (1-\beta)(-C_B)$, which implies $\beta \leq \frac{p-L_B}{K^G} \equiv \beta_3(p)$. Therefore, this pooling equilibrium in case 9 can be supported as a PBE if and only if $\beta \leq \beta_3(p)$.

Case 10: Consumer's beliefs are $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ and $\gamma \geq \overline{q}^B(p)$, that is, the green consumer buys regardless of the message she observes, and the brown consumer does not buy after observing a label, and does otherwise. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1-\beta)(-C_G - L_G) \geq \beta(p - C_G) + (1-\beta)(p - C_G)$, yielding $\beta \geq \frac{p+L_G}{p}$, which does not hold since $\beta \in [0,1]$ by definition; and the green firm has incentives to deviate towards no label. Hence, this pooling equilibrium in case 10 cannot be sustained as a PBE.

Case 11: Consumer's beliefs are $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ and $\overline{q}^G(p) \leq \gamma < \overline{q}^B(p)$, that is, the green consumer buys regardless of the message she observes, and the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1 - \beta)(-C_G - L_G) \geq \beta(p - C_G) + (1 - \beta)(-C_G)$, implying $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 11 cannot be supported as a PBE.

Case 12: Consumer's beliefs are $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ and $\gamma < \overline{q}^G(p)$, that is, the green consumer buys after observing a label, and she does not otherwise; and the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1 - \beta)(-C_G - L_G) \geq \beta(-C_G) + (1 - \beta)(-C_G)$, yielding $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1 - \beta)(-C_B - L_B) \geq \beta(-C_B) + (1 - \beta)(-C_B)$, which implies $\beta \geq \frac{L_B}{p - K^G} \equiv \beta_2(p)$. We know $\beta_1(p) \leq \beta_2(p)$. Hence, this pooling equilibrium in case 12 can be supported as a PBE if and only if $\beta \geq \beta_2(p)$ and $p \geq L_B + K^G$.

Price in Region C: Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\mu = q$ in equilibrium; and $\gamma \in [0, 1]$ off-the-equilibrium. After observing a label, the green consumer buys if and only if $qV_G^G + (1 - q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ since $V_B^G = 0$ by definition; and the brown consumer buys if and only if $qV_G^B + (1 - q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$. After observing no label, since the off-the-equilibrium beliefs satisfy $\gamma \in [0, 1]$, the green consumer buys if and only if $\gamma \geq \overline{q}^G(p)$; and the brown consumer buys if $\gamma \geq \overline{q}^B(p)$. Since $V_G^B < p$, $\overline{q}^B(p) > 1$; thus $q, \gamma \geq \overline{q}^B(p) \in [0, 1]$ is not satisfied, and the brown consumer does not buy regardless of the message she observes.

We analyze two different cases depending on the consumer's beliefs:

Case 1: Consumer's beliefs are $q, \gamma \geq \overline{q}^G(p)$, that is, the green consumer buys regardless of the message she observes; the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1 - \beta)(-C_G - L_G) \geq \beta(p - C_G) + (1 - \beta)(-C_G)$, yielding $L_G \leq 0$, which is not satisfied; and the green firm has incentives to deviate towards no label. Therefore, this pooling equilibrium in case 1 cannot be sustained as a PBE.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$ and $\gamma < \overline{q}^G(p)$, that is, the green consumer buys after observing a label, and she does not buy otherwise; the brown consumer does not buy regardless of the message she observes. Anticipating the consumers' response, the green firm acquires a label rather than no label if and only if $\beta(p - C_G - L_G) + (1 - \beta)(-C_G - L_G) \geq \beta(-C_G) + (1 - \beta)(-C_G)$, which is satisfied if $\beta \geq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B - L_B - K^G) + (1-\beta)(-C_B - L_B) \geq \beta(-C_B) + (1-\beta)(-C_B)$, which holds if $\beta \geq \frac{L_B}{p-K^G} \equiv \beta_2(p)$. Therefore, this pooling equilibrium in case 2 can be supported as a PBE if and only if $\beta \geq \beta_2(p)$ and $p \geq L_B + K^G$.

Intuitive Criterion (IC). In this pooling PBE, both types of firm acquire a label.

When p is in Region A (*case 1*), the green and the brown firm can obtain a higher payoff from deviating towards no label than their equilibrium payoff since $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)$ and $\beta(p - C_B) + (1-\beta)(p - C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)$, respectively. Hence, this pooling PBE in Proposition 3 (*case 1*) survives the Cho and Kreps' Intuitive Criterion.

When p lies in Region *B-I* (*cases 2a, 2b*) and in Region *B-II* (*cases 3a*, *3b*), the green and the brown firm can obtain a higher payoff from deviating towards no label than their equilibrium payoff since $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)$ and $\beta(p - C_B) + (1-\beta)(p - C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)$, respectively. Similarly, in cases *2c* and *3c*, the green and the brown firm also obtain a higher payoff from deviating. Therefore, since both types of firm have incentives to deviate towards no label, this pooling PBE in Proposition 3 (*case 2* and *case 3*) survives the Cho and Krep's Intuitive Criterion.

Finally, when p is in Region C, the green and the brown firm can obtain a lower payoff from deviating towards no label than equilibrium payoff since $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \geq \beta(-C_G - L_G) + (1-\beta)(-C_G - L_G)$ and $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \geq \beta(-C_B) + (1-\beta)(-C_B - L_B)$, respectively. Hence this pooling PBE in Proposition 3 (*case 3*) survives the Intuitive Criterion.

7.5 *Proof of Proposition 4*

Price in Region A**:** Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium; and $\mu \in [0,1]$ off-the-equilibrium. After observing no label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ since $V_B^G = 0$ by definition; and the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$. After observing a label, since the off-the-equilibrium beliefs satisfy $\mu \in [0,1]$, the green consumer buys if and only if $\mu \geq \overline{q}^G(p)$; and the brown consumer buys if $\mu \geq \overline{q}^B(p)$. Since $p - V_B^B < 0$ and $V_G^B - V_B^B > 0$, $\overline{q}^B(p) < 0$; thus $q, \mu \geq \overline{q}^B(p) \in [0,1]$ is always satisfied, and the brown consumer always buys regardless of the message she observes.

We analyze two different cases depending on the consumer's beliefs:

Case 1: Consumer's beliefs are $q, \mu \geq \overline{q}^G(p)$, that is, the green and the brown consumer buy regardless of the message they observe. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(p - C_B) \geq$

$\beta(p - C_B - L_B - K^G) + (1 - \beta)(p - C_B - L_B)$, implying $\beta \geq \frac{-L_B}{K^G}$, which holds since $\beta \in [0, 1]$ by definition. Therefore, this pooling equilibrium in case 1 can be sustained as a PBE.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$and $\mu < \overline{q}^G(p)$, that is, the green consumer buys after observing no label, and she does not otherwise; the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1 - \beta)(p - C_G) \geq \beta(-C_G - L_G) + (1 - \beta)(p - C_G - L_G)$, yielding $\beta \geq \frac{-L_G}{p}$, which holds. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1 - \beta)(p - C_B) \geq \beta(-C_B - L_B) + (1 - \beta)(p - C_B - L_B)$, implying $\beta \geq \frac{-L_B}{p}$, which is satisfied. Hence, this pooling equilibrium in case 2 can be supported as a PBE.

Price in Region B**:** Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium; and $\mu \in [0, 1]$ off-the-equilibrium. After observing no label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ since $V_B^G = 0$ by definition; and the brown consumer buys if and only if $qV_G^B + (1 - q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p - V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$. After observing a label, since the off-the-equilibrium beliefs satisfy $\mu \in [0, 1]$, the green consumer buys if and only if $\mu \geq \overline{q}^G(p)$; and the brown consumer buys if $\mu \geq \overline{q}^B(p)$.

We analyze 12 different cases depending on the consumer's beliefs:

Region *B-I*: $p \leq \widehat{p} \rightarrow \overline{q}^B(p) \leq \overline{q}^G(p)$:

Case 1: Consumer's beliefs are $q, \mu \geq \overline{q}^G(p)$, that is, the green and the brown consumer buy regardless of the message they observe. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1 - \beta)(p - C_G) \geq \beta(p - C_G - L_G) + (1 - \beta)(p - C_G - L_G)$, yielding $L_G \geq 0$, which is satisfied. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1 - \beta)(p - C_B) \geq \beta(p - C_B - L_B - K^G) + (1 - \beta)(p - C_B - L_B)$, implying $\beta \geq \frac{-L_B}{K^G}$, which holds since $\beta \in [0, 1]$ by definition. Therefore, this pooling equilibrium in case 1 can be sustained as a PBE.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$ and $\overline{q}^B(p) \leq \mu < \overline{q}^G(p)$, that is, the green consumer buys after observing no label, and does not otherwise; and the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1 - \beta)(p - C_G) \geq \beta(-C_G - L_G) + (1 - \beta)(p - C_G - L_G)$, implying $\beta \geq \frac{-L_G}{p}$, which holds. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1 - \beta)(p - C_B) \geq \beta(-C_B - L_B) + (1 - \beta)(p - C_B - L_B)$, yielding $\beta \geq \frac{-L_B}{p}$, which is satisfied. Hence, this pooling equilibrium in case 2 can be supported as a PBE.

Case 3: Consumer's beliefs are $q \geq \overline{q}^G(p)$ and $\mu < \overline{q}^B(p)$, that is, the green and the brown consumer buy after observing no label, and do not otherwise. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1 - \beta)(p - C_G) \geq \beta(-C_G - L_G) + (1 - \beta)(-C_G - L_G)$, yielding $p \geq -L_G$, which holds. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1 - \beta)(p - C_B) \geq \beta(-C_B - L_B) + (1 - \beta)(-C_B - L_B)$, implying $p \geq -L_B$, which is satisfied. Therefore, this pooling equilibrium in case 3 can be supported as a PBE.

Case 4: Consumer's beliefs are $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ and $\mu \geq \overline{q}^G(p)$, that is, the green consumer does not buy after observing no label, and does otherwise; and the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(-C_G)+(1-\beta)(p-C_G) \geq \beta(p-C_G-L_G)+(1-\beta)(p-C_G-L_G)$, which is satisfied if $\beta \leq \frac{L_G}{p} \equiv \beta_1(p)$. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(-C_B)+(1-\beta)(p-C_B) \geq \beta(p-C_B-L_B-K^G)+(1-\beta)(p-C_B-L_B)$, which holds if $\beta \leq \frac{L_B}{p-K^G} \equiv \beta_2(p)$. We know $\beta_1(p) \leq \beta_2(p)$. Hence, this pooling equilibrium in case 4 can be supported as a PBE if and only if $\beta \leq \beta_1(p)$.

Case 5: Consumer's beliefs are $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ and $\overline{q}^B(p) \leq \mu < \overline{q}^G(p)$, that is, the green consumer does not buy regardless of the message she observes, and the brown consumer buys regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(-C_G) + (1-\beta)(p-C_G) \geq \beta(-C_G-L_G) + (1-\beta)(p-C_G-L_G)$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(-C_B)+(1-\beta)(p-C_B) \geq \beta(-C_B-L_B) + (1-\beta)(p-C_B-L_B)$, implying $L_B \geq 0$, which holds. Therefore, this pooling equilibrium in case 5 can be supported as a PBE.

Case 6: Consumer's beliefs are $\overline{q}^B(p) \leq q < \overline{q}^G(p)$ and $\mu < \overline{q}^B(p)$, that is, the green consumer does not buy regardless of the message she observes, and the brown consumer buys after observing no label and does not otherwise. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(-C_G)+(1-\beta)(p-C_G) \geq \beta(-C_G-L_G)+(1-\beta)(-C_G-L_G)$, yielding $\beta \leq \frac{p+L_G}{p}$, which is satisfied since $\beta \in [0,1]$ by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(-C_B)+(1-\beta)(p-C_B) \geq \beta(-C_B-L_B)+(1-\beta)(-C_B-L_B)$, implying $\beta \leq \frac{p+L_B}{p}$, which holds. Hence, this pooling equilibrium in case 6 can be supported as a PBE.

Region *B-II*: $p > \widehat{p} \rightarrow \overline{q}^B(p) > \overline{q}^G(p)$:

Case 7: Consumer's beliefs are $q, \mu \geq \overline{q}^B(p)$, that is, the green and the brown consumer buy regardless of the message they observe. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p-C_G) + (1-\beta)(p-C_G) \geq \beta(p-C_G-L_G) + (1-\beta)(p-C_G-L_G)$, yielding $L_G \geq 0$, which is satisfied. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p-C_B) + (1-\beta)(p-C_B) \geq \beta(p-C_B-L_B-K^G) + (1-\beta)(p-C_B-L_B)$, implying $\beta \geq \frac{-L_B}{K^G}$, which holds since $\beta \in [0,1]$ by definition. Therefore, this pooling equilibrium in case 7 can be sustained as a PBE.

Case 8: Consumer's beliefs are $q \geq \overline{q}^B(p)$ and $\overline{q}^G(p) \leq \mu < \overline{q}^B(p)$, that is, the green consumer buys regardless of the message she observes; and the brown consumer buys after observing no label, and does not otherwise. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p-C_G)+(1-\beta)(p-C_G) \geq \beta(p-C_G-L_G)+(1-\beta)(-C_G-L_G)$, implying $\beta \leq \frac{p+L_G}{p}$, which holds. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p-C_B) + (1-\beta)(p-C_B) \geq \beta(p-C_B-L_B-K^G) + (1-\beta)(-C_B-L_B)$, yielding $\beta \leq \frac{p+L_B}{p}$, which is satisfied. Hence, this pooling equilibrium in case 8 can be supported as a PBE.

Case 9: Consumer's beliefs are $q \geq \overline{q}^B(p)$ and $\mu < \overline{q}^G(p)$, that is, the green and the brown consumer buy after observing no label, and do not otherwise. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(p - C_G) \geq \beta(-C_G - L_G) + (1-\beta)(-C_G - L_G)$, yielding $p \geq -L_G$, which holds. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(p - C_B) \geq \beta(-C_B - L_B) + (1-\beta)(-C_B - L_B)$, implying $p \geq -L_B$, which is satisfied. Therefore, this pooling equilibrium in case 9 can be supported as a PBE.

Case 10: Consumer's beliefs are $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ and $\mu \geq \overline{q}^B(p)$, that is, the green consumer buys regardless of the message she observes, and the brown consumer does not buy after observing no label, and does otherwise. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(-C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)$, which is satisfied if $\beta \geq \frac{p-L_G}{p} \equiv \beta_4(p)$. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(-C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)$, which holds if $\beta \geq \frac{p-L_B}{p+K^G} \equiv \beta_6(p)$. We check $\beta_4(p) \geq \beta_6(p)$ since $K^G(p - L_G) \geq p(L_G - L_B)$. Hence, this pooling equilibrium in case 10 can be supported as a PBE if and only if $\beta \geq \beta_4(p)$.

Case 11: Consumer's beliefs are $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ and $\overline{q}^G(p) \leq \mu < \overline{q}^B(p)$, that is, the green consumer buys regardless of the message she observes, and the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(-C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(-C_G - L_G)$, yielding $L_G \geq 0$, which is satisfied by definition. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(-C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(-C_B - L_B)$, implying $\beta \geq \frac{-L_B}{K^G}$, which holds since $\beta \in [0,1]$ by definition. Therefore, this pooling equilibrium in case 11 can be supported as a PBE.

Case 12: Consumer's beliefs are $\overline{q}^G(p) \leq q < \overline{q}^B(p)$ and $\mu < \overline{q}^G(p)$, that is, the green consumer buys after observing no label, and she does not otherwise; and the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(-C_G) \geq \beta(-C_G - L_G) + (1-\beta)(-C_G - L_G)$, yielding $\beta \geq \frac{-L_G}{p}$, which is satisfied. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(-C_B) \geq \beta(-C_B - L_B) + (1-\beta)(-C_B - L_B)$, implying $\beta \geq \frac{-L_B}{p}$, which holds. Hence, this pooling equilibrium in case 11 can be supported as a PBE.

Price in Region C: Upon observing the equilibrium message, the consumer cannot further update her beliefs about the firm's type, yielding $\gamma = q$ in equilibrium; and $\mu \in [0,1]$ off-the-equilibrium. After observing no label, the green consumer buys if and only if $qV_G^G + (1-q)V_B^G - p \geq 0$, which solving for q implies $q \geq \frac{p}{V_G^G} \equiv \overline{q}^G(p)$ since $V_B^G = 0$ by definition; and the brown consumer buys if and only if $qV_G^B + (1-q)V_B^B - p \geq 0$, which solving for q yields $q \geq \frac{p-V_B^B}{V_G^B - V_B^B} \equiv \overline{q}^B(p)$. After observing a label, since the off-the-equilibrium beliefs satisfy $\mu \in [0,1]$, the green consumer buys if and only if $\mu \geq \overline{q}^G(p)$; and the brown consumer buys if $\mu \geq \overline{q}^B(p)$. Since $V_G^B < p$, $\overline{q}^B(p) > 1$; thus $q, \mu \geq \overline{q}^B(p) \in [0,1]$ is not satisfied, and the brown consumer does not buy regardless of the message she observes.

We analyze two different cases depending on the consumer's beliefs:

Case 1: Consumer's beliefs are $q, \mu \geq \overline{q}^G(p)$, that is, the green consumer buys regardless of the message she observes, and the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather than acquiring a label if and only if $\beta(p - C_G) + (1-\beta)(-C_G) \geq \beta(p - C_G - L_G) + (1-\beta)(-C_G - L_G)$, yielding $L_G \geq 0$, which holds. Similarly, the brown firm does not label rather than acquiring a label if and only if $\beta(p - C_B) + (1-\beta)(-C_B) \geq \beta(p - C_B - L_B - K^G) + (1-\beta)(-C_B - L_B)$, implying $\beta \geq \frac{-L_G}{K^G}$, which is satisfied since $\beta \in [0,1]$ by definition. Therefore, this pooling equilibrium in case 1 can be sustained as a PBE.

Case 2: Consumer's beliefs are $q \geq \overline{q}^G(p)$and $\mu < \overline{q}^G(p)$, that is, the green consumer buys after observing no label, and she does not otherwise, and the brown consumer does not buy regardless of the message she observes. Anticipating the consumer's response, the green firm does not label rather acquiring a label if and only if $\beta(p-C_G)+(1-\beta)(-C_G) \geq \beta(-C_G-L_G)+(1-\beta)(-C_G-L_G)$, yielding $\beta \geq \frac{-L_G}{p}$, which holds. Similarly, the brown firm acquires a label rather than no label if and only if $\beta(p - C_B) + (1-\beta)(-C_B) \geq \beta(-C_B - L_B) + (1-\beta)(-C_B - L_B)$, yielding $\beta \geq \frac{-L_G}{p}$, which is satisfied since $\beta \in [0,1]$ by definition. Hence, this pooling equilibrium in case 2 can be supported as a PBE.

Intuitive Criterion (IC). In this pooling PBE, both firms do not label.

When p is in Region A (*case 1*), both types of consumer buy after observing no label. Then, the green and the brown firm can obtain a lower payoff from deviating towards a label than their equilibrium payoff since $\beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G) \leq \beta(p - C_G) + (1-\beta)(p - C_G)$ and $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \leq \beta(p - C_B) + (1-\beta)(p - C_B)$, respectively. In summary, no type of firm has incentives to deviate towards a label, entailing that the consumer cannot further update her off-the-equilibrium beliefs $\mu \in [0,1]$. Therefore, this pooling PBE in Proposition 4 (*case 1*) survives the Cho and Kreps' Intuitive Criterion.

When p lies in Region *B-I* (*case 2a*) and in Region *B-II* (*case 3a*), both types of consumer buy after observing no label. The green and the brown firm can obtain a lower payoff from deviating towards a label than their equilibrium payoff since $\beta(p-C_G-L_G)+(1-\beta)(p-C_G-L_G) \leq \beta(p-C_G)+(1-\beta)(p-C_G)$ and $\beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B) \leq \beta(p - C_B) + (1-\beta)(p - C_B)$, respectively. Similarly, in *cases 2b* and *3b* (in which the green (brown) consumer buys and the brown(green) does not buy, respectively), the green and the brown firm also obtain a lower payoff from deviating towards a label if equilibrium conditions $\beta \leq \beta_1(p)$ in *case 2b*, and $\beta \geq \beta_4(p)$ in *case 3b*, are satisfied. Hence, this pooling PBE in Proposition 4 (*case 2* and *case 3*) survives the Intuitive Criterion.

When p is in Region C (*case 4*), the green consumer buys while the brown consumer does not buy after observing a label. The green and the brown firm can obtain a lower payoff from deviating towards a label than equilibrium payoff since $\beta(p-C_G-L_G)+(1-\beta)(-C_G-L_G) \leq \beta(p-C_G)+(1-\beta)(-C_G)$ and $\beta(p-C_B-L_B-K^G)+(1-\beta)(-C_B-L_B) \leq \beta(p-C_B)+(1-\beta)(-C_B-L_B)$, respectively. Therefore, this pooling PBE in Proposition 4 (*case 4*) survives the Intuitive Criterion.

7.6 *Proof of Corollary 1*

Same cost of labeling for both type of firm, $L_B = L_G$. When the labeling cost of the brown firm coincides with that of the green firm, the SE of Proposition 1 can be sustained under more restrictive conditions on β for all regions *A*-*C*. In particular, cutoff $\beta_1(p) \equiv \frac{L_G}{p}$ is unaffected; cutoff $\beta_2(p)$ becomes $\beta_2(p) \equiv \frac{L_G}{p-K^G}$, which is lower than when allowing for different labeling costs, $\frac{L_B}{p-K^G}$; and cutoff $\beta_3(p)$ becomes $\beta_3(p) \equiv \frac{p-L_G}{K^G}$ which is higher than when allowing for different labeling costs, $\frac{p-L_B}{K^G}$. As a consequence, the condition on β that supports regions A and C, $\beta_1(p) \leq \beta \leq \beta_2(p)$, becomes more demanding, and a similar argument applies to the condition on β sustaining region B, $\beta \geq \beta_3(p)$.

In the case of the PE of Proposition 3, assuming that both firms face the same labeling cost implies that this equilibrium can be sustained under larger conditions on β. Specifically, region A, which can be sustained if $\beta \geq \beta_2(p)$, can now be supported under larger conditions since cutoff $\beta_2(p)$ is lower as discussed above. A similar analysis applies to region *B-I* since cutoff $\beta_2(p)$ is lower, while cutoffs $\beta_3(p)$ and $\beta_5(p)$ are higher, thus expanding the range of β that sustains this equilibrium. This argument also applies to region *B-II* and C given that, respectively, cutoffs $\beta_6(p)$ and $\beta_3(p)$ are higher and cutoff $\beta_2(p)$ is lower.

Same greenwashing penalty from both types of consumers, $K^G = 0$. When both types of consumers penalize greenwashing practices in the same way, $k^G = k^B$, entailing $K^G = 0$. Hence, the SE of Proposition 1 is supported under more restrictive conditions on β for all regions *A*-*C*. In particular, cutoff $\beta_1(p) \equiv \frac{L_G}{p}$ is unaffected; cutoff $\beta_2(p)$ becomes $\beta_2(p) \equiv \frac{L_B}{p}$, which is lower than when allowing for asymmetric greenwashing penalties, $\frac{L_B}{p-K^G}$; and cutoff $\beta_3(p)$ approaches infinity, thus being higher than when allowing for different greenwashing penalties, $\frac{p-L_B}{K^G}$. As a consequence, the condition on β that supports regions A and C, $\beta_1(p) \leq \beta \leq \beta_2(p)$, becomes more demanding, and a similar argument applies to the condition on β sustaining region B, $\beta \geq \beta_3(p)$, which cannot be supported in this context.

In the case of the PE of Proposition 3, assuming that both firms face the same greenwashing penalty implies that this equilibrium can be sustained under larger conditions on β. Specifically, region A, which can be sustained if $\beta \geq \beta_2(p)$, can now be supported under larger conditions since cutoff $\beta_2(p)$ is lower as discussed above. A similar analysis applies to region *B-I* since cutoff $\beta_2(p)$ is lower, while cutoff $\beta_3(p)$ is higher and $\beta_5(p)$ is unaffected, thus weakly expanding the range of β that sustains this equilibrium. This argument also applies to region *B-II* and C given that, respectively, cutoffs $\beta_6(p)$ and $\beta_3(p)$ are higher and cutoff $\beta_2(p)$ is lower.

7.7 *Proof of Corollary 4*

We first compare each cutoff β against each of the other cutoffs.

Comparing $\beta_1(p)$ and $\beta_2(p)$, we find that $\beta_1(p) \le \beta_2(p)$ implies $p(L_G - L_B) \le L_G K^G$, which is satisfied since $L_B \ge L_G$ by definition. In addition, $\beta_1(p) \le \beta_3(p)$ yields $L_G K^G \le p(p - L_B)$, which holds if $p \ge L_B + K^G$. Furthermore, $\beta_1(p) \le \beta_4(p)$ implies $L_G \le p - L_G$, which is satisfied if $p \ge 2L_G$. Comparing $\beta_1(p)$ and $\beta_5(p)$, we find that $\beta_1(p) \le \beta_5(p)$ yields $L_G \le p - L_B$, which holds if $p \ge L_B + L_G$. Finally, $\beta_1(p) \le \beta_6(p)$ implies $p(p - L_B - L_G) \ge L_G K^G$, which is satisfied if $p \ge L_B + L_G + K^G$.

We now compare cutoff $\beta_2(p)$ against other cutoffs. First, $\beta_2(p) \le \beta_3(p)$ yields $L_B K^G \le (p - K^G)(p - L_B)$, which holds if $p \ge K^G$. Second, $\beta_2(p) \le \beta_4(p)$ implies $L_B p \le (p - K^G)(p - L_G)$, which is satisfied if $p \ge K^G$. Third, $\beta_2(p) \le \beta_5(p)$ yields $L_B p \le (p - K^G)(p - L_B)$, which is satisfied if $p \ge K^G$. Finally, $\beta_2(p) \le \beta_6(p)$ implies $L_B(p + K^G) \le (p - K^G)(p - L_B)$, which is satisfied if $p \ge K^G$.

We now compare cutoff $\beta_3(p)$ against other cutoffs. First, we obtain that $\beta_3(p) \ge \beta_4(p)$ yields $p(p - L_B) \le K^G(p - L_G)$, which holds if $p \ge K^G$. Second, $\beta_3(p) \ge \beta_5(p)$ implies $p(p - L_B) \le K^G(p - L_B)$, which is satisfied if $p \ge K^G$. Finally, $\beta_3(p) > \beta_6(p)$ holds since $K^G < p + K^G$.

We now rank cutoff $\beta_4(p)$ against other cutoffs. First, $\beta_4(p) \ge \beta_5(p)$ since $L_G \le L_B$ by definition. Second, $\beta_4(p) \ge \beta_6(p)$ holds since $(p - L_G) \ge (p - L_B)$ and if $p \le p + K^G$, which means $K^G \ge 0$.

Finally, $\beta_5(p) \ge \beta_6(p)$ is satisfied if $p \le p + K^G$, which means $K^G \ge 0$.

We next rank cutoffs $\beta_1(p)$ to $\beta_6(p)$:

- When $p \ge L_B + K^G$, we know that $\beta_3(p) > 1$. In addition, from above comparisons, we know that $\beta_1(p) \le \beta_2(p)$, $\beta_6(p) \le \beta_5(p) \le \beta_4(p)$, and $\beta_2(p) \le \beta_6(p)$ if $p \ge K^G$. Given the condition $p \ge L_B + K^G$, $\beta_2(p) \le \beta_6(p)$ holds since the condition for the opposite, $p < K^G$, does not hold. Therefore, if $p \ge L_B + K^G$, cutoffs are ranked as follows $0 < \beta_1(p) \le \beta_2(p) \le \beta_6(p) \le \beta_5(p) \le \beta_4(p) < 1 < \beta_3(p)$.

- When $p < L_B + K^G$, we know that $\beta_2(p) > 1$. In addition, from above comparisons, we know that if this condition holds then $\beta_1(p) > \beta_3(p)$. In addition, we know that $\beta_3(p) > \beta_6(p)$ and $\beta_6(p) \le \beta_5(p) \le \beta_4(p)$.

  1. If $p \ge K^G$, $\beta_3(p) \ge \beta_4(p)$, and the cutoff ranking becomes $0 < \beta_6(p) \le \beta_5(p) \le \beta_4(p) \le \beta_3(p) < \beta_1(p) < 1 < \beta_2(p)$.
  2. If $2L_G \ge K^G > p$, $\beta_3(p) < \beta_5(p)$ and $\beta_1(p) > \beta_4(p)$. The cutoff ranking becomes $0 < \beta_6(p) < \beta_3(p) < \beta_5(p) \le \beta_4(p) < \beta_1(p) < 1 < \beta_2(p)$.
  3. If $L_B + L_G \ge K^G > \max\{p, 2L_G\}$, $\beta_3(p) < \beta_5(p)$ and $\beta_5(p) < \beta_1(p) \le \beta_4(p)$. The cutoff ranking becomes $0 < \beta_6(p) < \beta_3(p) < \beta_5(p) < \beta_1(p) \le \beta_4(p) < 1 < \beta_2(p)$.

4. If $K^G > \max\{p, L_B + L_G\}$, $\beta_3(p) < \beta_5(p)$ and $\beta_1(p) \leq \beta_5(p)$. The cutoff ranking becomes $0 < \beta_6(p) < \beta_3(p) < \beta_1(p) \leq \beta_5(p) \leq \beta_4(p) < 1 < \beta_2(p)$.

7.8 *Proof of Corollary 5*

Price in Region A**:** When $p < L_B + K^G$, only a SE can arise if β satisfies $\beta \geq \beta_1(p)$. When $p \geq L_B + K^G$, the SE in Proposition 1.a (which we denoted SE1a) arises if β satisfies $\beta_1 \leq \beta \leq \beta_2$, and the PE in Proposition 3.1 (denoted as PE1) can be supported as a PBE if $\beta \geq \beta_2(p)$. The conditions on β supporting SE1a and PE1 do not allow them to coexist.

Price in Region B-I**:** When $p < L_B + K^G$, the SE in Proposition 1.b (denoted as SE1b) cannot coexist with the PE in Proposition 3.2b (identified as PE2b) because the condition on β makes them incompatible (i.e., $\beta \geq \beta_3$ and $\beta < \beta_3$, respectively). However, SE1b can coexist with the PE in Proposition 3.2c (denoted as PE2c) when $\beta \in [\beta_3(p), \beta_5(p)]$ if and only if $p < K^G$. In addition, PE2b can also coexist with PE2c if $\beta \leq \min\{\beta_1(p), \beta_5(p)\}$.

When $p \geq L_B + K^G$, SE1b cannot be sustained. However, the three PEs in Proposition 3.2a, 3.2b, and 3.2c (PE2a, PE2b, and PE3c, respectively) can coexist if β satisfies $\beta \in [\beta_2(p), \beta_5(p)]$.

Price in Region B-II: When $p < L_B + K^G$, the separating PBE SE1b cannot coexist with the PEs in Proposition 3.3a and 3.3b (denoted as PE3b and PE3c, respectively) because the incompatibility on the β conditions. However, PE3a and PE3b can coexist if $\beta \leq \beta_6(p)$.

When $p \geq L_B + K^G$, SE1b cannot be supported. However, PE3a and PE3b can coexist with the PE in Proposition 3.3c (denoted as PE3c) if β satisfies $\beta \in [\beta_2(p), \beta_6(p)]$.

Price in Region C: When $p < L_B + K^G$ only a separating PBE can arise if $\beta \geq \beta_1(p)$. When $p \geq L_B + K^G$, the SE in Proposition 1.2 (denoted as SE2) arises if β satisfies $\beta_1(p) \leq \beta \leq \beta_2(p)$, and the PE in Proposition 3.4 (denoted as PE4) can be supported if $\beta \geq \beta_2(p)$. The condition of β for the existence of SE2 and PE4, however, prevents their coexistence.

7.9 *Proof of Proposition 5*

Price in Region A**:** When $p < L_B + K^G$, only a SE can arise if β satisfies $\beta \geq \beta_1(p)$. When $p \geq L_B + K^G$, SE1a and PE1 cannot coexist, as shown in the proof of Corollary 5. Therefore, no welfare comparison is required.

Price in Region B-I**:** When $p < L_B + K^G$, SE1b can coexist with PE2c. Comparing their welfare levels, we obtain that, when the firm is green,

$$\begin{aligned} SW^G_{SE1b} - SW^G_{PE2c} &= [\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \\ &\quad -[\beta \times 0 + (1-\beta)(V^B_G - p) + \beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G)], \end{aligned}$$

which simplifies to $\beta V^G_G \geq 0$. Hence, SE1b generates a higher expected welfare than PE2c when the firm is green. When the firm is brown, we find

$$\begin{aligned} SW^B_{SE1b} - SW^B_{PE2c} &= [\beta \times 0 + (1-\beta) \times 0 + \beta(-C_B) + (1-\beta)(-C_B)] \\ &\quad -[\beta \times 0 + (1-\beta)(V^B_B - p) + \beta(-C_B - L_B) + (1-\beta)(p - C_B - L_B)], \end{aligned}$$

which simplifies to $-V^B_B + \beta V^B_B + L_B$. Therefore, under bilateral uncertainty (since the social planner does observe the firm's type), we have that the expected change in social welfare is

$$q[SW^G_{SE1b} - SW^G_{PE2c}] + (1-q)[SW^B_{SE1b} - SW^B_{PE2c}] = q[\beta V^G_G] + (1-q)[-V^B_B + \beta V^B_B + L_B],$$

which is positive if and only if

$$\frac{q}{1-q} \geq \frac{(1-\beta)V^B_B - L_B}{\beta V^G_G}.$$

Therefore, SE1b generates more expected welfare than PE2c when these equilibria coexist (i.e., $p < L_B + K^G$ and $\beta \in [\beta_3(p), \beta_5(p)]$ and $p < K^G$) if and only if q satisfies $\frac{q}{1-q} \geq \frac{(1-\beta)V^B_B - L_B}{\beta V^G_G}$.

Still in the context where $p < L_B + K^G$, we know from Corollary 5, that PE2b can coexist with PE2c if $\beta \leq \min\{\beta_1(p), \beta_5(p)\}$. Comparing their welfare levels, we obtain that, when the firm is green,

$$\begin{aligned} SW^G_{PE2b} - SW^G_{PE2c} &= [\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \\ &\quad -[\beta \times 0 + (1-\beta)(V^B_G - p) + \beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G)] \end{aligned}$$

which simplifies to $\beta V^G_G \geq 0$. Therefore, PE2b generates more expected welfare than PE2c when the firm is green. Similarly, when the firm is brown, we have

$$\begin{aligned} SW^B_{PE2b} - SW^B_{PE2c} &= [\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \\ &\quad -[\beta \times 0 + (1-\beta)(V^B_B - p) + \beta(-C_B - L_B) + (1-\beta)(p - C_B - L_B)] \end{aligned}$$

which reduces to $\beta[V^G_B - K^G]$, and this result simplifies to $-\beta K^G \leq 0$ given that $V^G_B = 0$. Hence, PE2b yields a lower welfare than PE2c when the firm is brown. Therefore, under bilateral uncertainty, we obtain

$$q[SW^G_{PE2b} - SW^G_{PE2c}] + (1-q)[SW^B_{PE2b} - SW^B_{PE2c}] = q\left[\beta V^G_G\right] + (1-q)[-\beta K^G],$$

which is positive if and only if

$$\frac{q}{1-q} \geq \frac{K^G}{V^G_G}.$$

Then, PE2b yields a higher expected welfare than PE2c when these equilibria coexist (i.e., $p < L_B + K^G$ and $\beta \leq \min\{\beta_1(p), \beta_5(p)\}$) if and only if q satisfies $\frac{q}{1-q} \geq \frac{K^G}{V^G_G}$

When $p \geq L_B + K^G$, we know from Corollary 5 that only PE2a, PE2b, and PE3c can coexist if $\beta \in [\beta_2(p), \beta_5(p)]$. We first compare the welfare from PE2a against PE2b when the firm is green, obtaining that

$$\begin{aligned} SW^G_{PE2a} - SW^G_{PE2b} &= [\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \\ &\quad -[\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \end{aligned}$$

which collapses to zero. Similarly, when the firm is brown, the change in expected social welfare is

$$\begin{aligned} SW^B_{PE2a} - SW^B_{PE2b} &= [\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \\ &\quad -[\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \end{aligned}$$

which also reduces to zero. Therefore, under bilateral uncertainty, we have

$$q[SW^G_{PE2a} - SW^G_{PE2b}] + (1-q)[SW^B_{PE2a} - SW^B_{PE2b}] = 0.$$

implying that PE2a and PE2b generate the same expected social welfare when they coexist, i.e., $p \geq L_B + K^G$ and $\beta \in [\beta_2(p), \beta_5(p)]$.

Still in a setting where $p \geq L_B + K^G$, we next compare the welfare of PE2a (or PE2b, since they both yield the same welfare level) against that of PE3c. When the firm is green, we obtain

$$\begin{aligned} SW^G_{PE2a=PE2b} - SW^G_{PE2c} &= [\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \\ &\quad -[\beta \times 0 + (1-\beta)(V^B_G - p) + \beta(-C_G - L_G) + (1-\beta)(p - C_G - L_G)] \end{aligned}$$

which simplifies to $\beta V^G_G \geq 0$. Similarly, when the firm is brown, we find

$$\begin{aligned} SW^B_{PE2a=PE2b} - SW^B_{PE2c} &= [\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \\ &\quad -[\beta \times 0 + (1-\beta)(V^B_B - p) + \beta(-C_B - L_B) + (1-\beta)(p - C_B - L_B)] \end{aligned}$$

which reduces to $\beta[V^G_B - K^G]$, further simplifying to $-\beta K^G \leq 0$ given that $V^G_B = 0$. Therefore, under bilateral uncertainty, we have

$$q[SW^G_{PE2a=PE2b} - SW^G_{PE2c}] + (1-q)[SW^B_{PE2a=PE2b} - SW^B_{PE2c}] = q\left[\beta V^G_G\right] + (1-q)[-\beta K^G]$$

which is positive if and only if

$$\frac{q}{1-q} \geq \frac{K^G}{V^G_G}.$$

Then, PE2a and PE2b generate a higher welfare than PE2c when they coexist (i.e., $p \geq L_B + K^G$ and $\beta \in [\beta_2(p), \beta_5(p)]$) if and only if $\frac{q}{1-q} \geq \frac{K^G}{V^G_G}$.

Price in Region *B-II*: When $p < L_B + K^G$, we know from Corollary 5 that only PE3a and

PE3b can coexist. Comparing their welfare levels, we find that, when the firm is green

$$\begin{aligned} SW^G_{PE3a} - SW^G_{PE3b} &= [\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \\ &\quad -[\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \end{aligned}$$

which collapses to zero. Similarly, when the firm is brown, we have that

$$\begin{aligned} SW^B_{PE3a} - SW^B_{PE3b} &= [\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \\ &\quad -[\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \end{aligned}$$

which also reduces to zero. Therefore, under bilateral uncertainty, we find that

$$q[SW^G_{PE3a} - SW^G_{PE3b}] + (1-q)[SW^B_{PE3a} - SW^B_{PE2b}] = 0,$$

implying that both PE3a and PE3b generate the same welfare when they coexist, i.e., $p < L_B + K^G$ and $\beta \leq \beta_6(p)$.

When $p \geq L_B + K^G$, Corollary 5 showed that three PEs coexist: PE3a, PE3b, and PE3c. From our above discussion, we know that PE3a and PE3b generate the same welfare level. We can now compare the welfare under PE3a and PE3c obtaining that, when the firm is green,

$$\begin{aligned} SW^G_{PE3a=PE3b} - SW^G_{PE3c} &= [\beta(V^G_G - p) + (1-\beta)(V^B_G - p) + \beta(p - C_G - L_G) + (1-\beta)(p - C_G - L_G)] \\ &\quad -[\beta(V^G_G - p) + (1-\beta) \times 0 + \beta(p - C_G - L_G) + (1-\beta)(-C_G - L_G)] \end{aligned}$$

which simplifies to $(1-\beta)V^B_G \geq 0$. Similarly, when the firm is brown, we find that

$$\begin{aligned} SW^B_{PE3a=PE3b} - SW^B_{PE3c} &= [\beta(V^G_B - p) + (1-\beta)(V^B_B - p) + \beta(p - C_B - L_B - K^G) + (1-\beta)(p - C_B - L_B)] \\ &\quad -[\beta(V^G_B - p) + (1-\beta) \times 0 + \beta(p - C_B - L_B - K^G) + (1-\beta)(-C_B - L_B)] \end{aligned}$$

which reduces to $(1-\beta)V^B_B \geq 0$. Therefore, under bilateral uncertainty, PE3a and PE3b yield an unambiguous highest expected welfare than PE3c when these equilibria coexist, i.e., $p \geq L_B + K^G$ and $\beta \in [\beta_2(p), \beta_6(p)]$.

Price in Region C: When $p < L_B + K^G$, only a SE can arise if $\beta \geq \beta_1(p)$. When $p \geq L_B + K^G$, the conditions on β for SE2 and PE4 to arise are incompatible. As a consequence, no welfare comparison is required.

References

[1] Arguedas, C. and Blanco, E. (2014) "On fraud and certification of corporate social responsibility." *Working papers in Economics and Statistics*, 2014-18, University of Innsbruck, Research

Platform Empirical and Experimental Economics (eeecon), Innsbbruck.

[2] Bachmann, R., Ehrlich, G., Fan, Y., and Ruzic, D. (2019) "Firms and collective reputation: A study of the Volkswagen emission scandal." *NBER Working Paper Series*, 26117.

[3] Baksi, S., and P. Bose (2007) "Credence Goods, Efficient Labelling Policies, and Regulatory Enforcement." *Environmental and Resource Economics*, 37(2):411–430.

[4] Berrone, P., Fosfuri, A., and Gelabert, L. (2017) "Does Greenwashing Pay Off? Understanding the Relationship Between Environmental Actions and Environmental Legitimacy." *Journal of Business Ethics*, 144:363-379.

[5] Bottega, L. and De Freitas, J. (2019) "Imperfect certification in a Bertrand duopoly." *Economics Letters*, 178:33-36.

[6] Chatterje, K. and Samuelson, L. (1983) "Bargaining under Two-Sided Incomplete Information: The Unrestricted Offers Case." *Operations Research*, 36(4):605-618.

[7] Cho, I.K. and Kreps, D.M. (1987) "Signaling Games and Stable Equilibria." *Quarterly Journal of Economics,* 102(2):179-222.

[8] Cone Communications (2013) *Green Gap Trend Tracker*. Available at: http://www.conecomm.com/research-blog/2013-cone-communications-green-gap-trend-tracker (accessed Mar 4, 2019)

[9] Connelly, B.L., Certo, S.T., Ireland, R.D., and Reutzel, C.R. (2011) "Signaling Theory: A Review and Assessment." *Journal of Management,* 37(1):39-67.

[10] Cramton, P.C. (1992) "Strategic Delay in Bargaining with Two-Sided Uncertainty." *The Review of the Economic Studies*, 59(1):205-225.

[11] Daughety, A. F., and J. F. Reinganum (2008) "Communicating Quality: A Unified Model of Disclosure and Signalling." *RAND Journal of Economics*, 39:973–989.

[12] Delmas, M.A. and Burbano, V.C. (2011) "The Drivers of Greenwashing." *California Management Review*, 54(1):64-87.

[13] Ecolabel Index (2019). Available at: http://www.ecolabelindex.com/ (accessed March 15, 2019)

[14] European Commission (2014) *Consumer Market Study on Environmental Claims for Non-Food Products*. Directorate-General for Justice and Consumers. Available at: https://ec.europa.eu/info/sites/info/files/study_on_environmetal_claims_for_non_food_products_2014_

[15] Gale, D. (2001) "Signaling in markets with two-sided adverse selection." *Economic Theory*, 18:391-414.

[16] Garrido, D., Espinola-Arredondo, A., and Munoz-Garcia, F. (2020) "Can Mandatory Certification Promote Greenwashing? A Signaling Approach." Journal of Public Economic Theory, 22(6):1801–1851.

[17] Gruere, G. (2013) "A Characterization of Environmental Labelling and Information Schemes." *OECD Environment Working Papers*, 2.

[18] Hamilton, S.F. and Zilberman, D. (2006) "Green markets, eco-certification, and equilibrium fraud." *Journal of Environmental Economics and Management*, 52:627-644.

[19] Ibanez, L., and G. Grolleau (2008) "Can Ecolabeling Schemes Preserve the Environment?" *Environmental and Resource Economics*, 40:233–249.

[20] Kollewe, J. (2016) "VW profits down 20% after the diesel emission scandal." *The Guardian*, May 31, 2016. Available at: https://www.theguardian.com/business/2016/may/31/vw-volkswagen-profits-down-20-diesel-emissions-scandal (accessed September 23, 2019)

[21] Lyon, T.P. and Maswell, J.W. (2011) "Greenwash: Corporate Environmental Disclosure under Threat of Audit." *Journal of Economics and Management Strategy*, 20(1):3-41.

[22] Lyon, T.P. and Montgomery, A.W. (2015) "The Means and End of Greenwash." *Organization & Environment*, 28(2):223-249.

[23] Mahenc, P. (2008) "Signaling the environmental performance of polluting products to green consumers." *International Journal of Industrial Organization*, 26:59-68.

[24] Mahenc, P. (2017) "Honest versus misleading certification." *Journal of Economics & Management Strategy*, 26(2):454-483.

[25] Mason, C.F. (2011) "Eco-Labeling and Market Equilibrium with Noisy Certification Tests." *Environmental Resource Economics*, 48:537-560.

[26] Nielsen (2018) "Was 2018 the Year of the Influential Sustainable Consumer?." *CPG, FMCG & Retail Article,* 12-17-2018. Available at: https://www.nielsen.com/us/en/insights/news/2018/was-2018-the-year-of-the-influential-sustainable-consumer.html (accessed March 4, 2019)

[27] Reputation Institute (2016) *Case Study: Volkswagen. How VW Lost the Public's Trust*. Available at: https://www.reputationinstitute.com/research/how-vw-lost-publics-trust (accessed August 27, 2019)

[28] Terrachoice (2010) *The sins of greenwashing. Home and family edition 2010. A report of environmental claims made in North America consumer market*. Northbrook, IL: Underwriters Laboratories.

[29] Volle, A. (2017) “Why is price useless to signal environmental quality?.” *FAERE Working Paper*, 2017.30.

[30] Yokessa, M. and Marette, S. (2019) "A Review of Econ-labels and their Economic Impact." *International Review of Environmental and Resource Economics*, 13:119-163.