On Fraud and Certification of Corporate Social Responsibility*

Carmen Arguedas[a] and Esther Blanco[b,c]

[a] Departamento de Análisis Económico: Teoría Económica e Historia Económica, Universidad Autónoma de Madrid.

[b] Department of Public Finance, University of Innsbruck;

[c]Ostrom Workshop in Political Theory and Policy Analysis, Indiana University.

June 5, 2014

Abstract

We analyze the strategic decision of firms to voluntarily certify corporate social responsibility (CSR) practices in a context where other firms can falsely pretend to be socially responsible. Equilibrium outcomes are crucially determined by consumers' beliefs about the credibility of firms' CSR claims, which depend in turn on the (expected) fines for fraud. First, we show that an increase in such fines extends the likelihood of firms investing in CSR, at the expense of a reduced likelihood of certification. Second, fraud only arises when the fines for fraud are at intermediate levels and some CSR firms do not certify their practices. Third, the presence of fraud comes at a cost for firms by inducing lower equilibrium prices than in settings with honest marketing. Forth, the coexistence of fraud and certification induces differentiation price premia below marginal production costs and certification price premia above marginal certification costs. Lastly, social welfare rises as fines for fraud increase.

Key words: corporate social responsibility, credence goods, certification, fraud.

JEL codes: C72, D43, H23, Q58.

*Please send all correspondence to: Carmen Arguedas, Departamento de Análisis Económico: Teoría Económica e Historia Económica, Universidad Autónoma de Madrid, 28049 Cantoblanco, Madrid, Spain. E-mail: carmen.arguedas@uam.es. The authors wish to thank Rudolf Kerschbamer and the participants of the 2013 EAERE Annual Meeting, the 38th Conference of the Spanish Economic Association and the 2013 Annual Meeting of the Austrian Economic Association for useful comments and suggestions on earlier drafts. Carmen Arguedas acknowledges the financial support from the Spanish Government under research project ECO2011-25349.

1. Introduction

Corporate social responsibility (CSR) has become a mainstream business activity (Kitzmueller and Shimshack, 2012). It includes voluntary investments by firms to over-comply with social and environmental regulations that result often times in the provision of credence goods. The credence goods associated to CSR are product attributes difficult to judge even after purchase (Dulleck et al., 2011), and include characteristics of production, such as employment discrimination, "fair trade", and human rights (Baksi and Bose, 2007; Feddersen and Gilligan, 2001), or environmental aspects of products that consumers cannot observe directly, such as "Dolphin safe" tuna or genetically modified organism (Baksi and Bose, 2007; Harbaugh et al., 2011). This results in an asymmetric information problem between firms and consumers whereby miss-leading marketing (fraud) becomes an option (see Lyon and Maxwell, 2011 for a theoretical model of greenwash). Given the existence of fraud, firms' credibility among consumers is questioned, and makes certification an option to alleviate the asymmetric information problem. In fact, the use of certification of CSR is widespread; more than one third of large firms have voluntary external certifications for social and environmental standards (Kitzmueller and Shimshack, 2012). As an example, the Ecolabel index (www.ecolabelindex.com) is currently tracking 448 ecolabels in 197 countries, and 25 industry sectors.

Interestingly, still, many other firms base their CSR communication to the public on news releases and marketing campaigns without third-party validation. This is the case despite scandals for fraudulent marketing among non-certified firms are not rare. Recurrent examples are the advertisement as "sustainable energy sources" of major oil infrastructure (e.g. refineries) or extraction projects (e.g. oil platforms). The fact that firms continue to invest in marketing their CSR actions without certification suggests

that such claims have some information content for consumers that are worth the marketing investments. Empirical findings seem to support this view. For example, up to almost half of EU consumers self-report to trust completely or rather trust producers' claims on their products' environmental performance (Commission, 2009).

In sum, the coexistence of conventional (non-CSR) production, CSR practices, fraud and certification of CSR is present in many industries. To our knowledge, the combination of these issues has not been explored in the existing literature up to date, although we believe it has key effects on the likelihood of CSR practices and the associated mark-ups. In this context, we particularly focus on the role of the (expected) fines for fraud and their associated influence on consumers' trust on firms' claims. We show how these two related aspects affect (i) the likelihood of CSR practices and that of subsequent certification; (ii) the likelihood of fraud in the industry, and (iii) firms' profits and social welfare.

In terms of modeling, we treat firms engaging in CSR practices as producers of a high quality version of a product, while firms that use conventional practices are assumed to provide a standard (or low quality) version. We present an oligopoly model of vertical product differentiation where all consumers prefer the high- over the low-quality version of the product for the same price.[1] We assume that product quality is not directly observable by the consumers. As a result, firms can commit fraud about product quality and consumers form beliefs about the likelihood of honest behavior.

Our model includes a broader strategy set for firms than the one usually considered in the literature, by combining two interacting dimensions. The first

[1] Khanna (2001), Lyon and Maxwell (2008), Portney (2008), and more recently Kitzmueller and Shimshack (2012), present excellent reviews of the literature on motivations of firms to undertake CSR. In addition to the demand-driven efforts, which is the focus of this study, these papers include other motivations, such as incentives in employment contracts, "warm glow" preferences, social norms, and private (NGOs) and public politics.

dimension embraces the *production technology*, since firms can decide to produce either the high or the low quality version of the product. The second dimension addresses the *marketing* strategy. Firms producing the low quality version can fraudulently (truthfully) market their product to consumers as high (low) quality. In addition, firms producing the high quality version can decide to provide (or not) third-party validation through certification. Consumers' equilibrium beliefs with regards to the validity of firms' uncertified high-quality claims is the key element in the analysis, which crucially depends on the costs of producing high quality and the (expected) fines for fraud.

We derive three main sets of results. The first has to do with the influence of the fines for fraud on the likelihood of CSR (high quality) production and subsequent certification. Everything else equal, increasing fines for fraud induces an increased likelihood of CSR production. Intuitively, sufficiently large fines deter firms producing the conventional product to commit fraud. This induces consumers to (consistently) believe that the uncertified firms that claim to offer CSR do so truthfully. The increased likelihood of CSR production comes at the cost of a decreased likelihood of certification. Given that consumers trust firms' uncertified claims, certification has no influence in demand, as it does not provide any additional information. Conversely, widespread certification of CSR investments arises under low fines for fraud, where consumers do not trust firms' claims at all.

Second, we link fines for fraud with the likelihood of fraud in the industry. We show that fraud about firms' quality can only arise in settings with intermediate fines and where the conventional and the uncertified CSR versions coexist in the market. The key aspect is consumers' imperfect trust on the uncertified variety. At the extreme case of large fines fraud does not arise, since fines are deterrent. But, surprisingly, fraud does not arise under low fines either. In this alternative scenario, firms producing the

conventional product are obviously tempted to cheat, but they anticipate that by doing so consumers will (consistently) believe that they are fraudulent. As a result, firms truthfully offering CSR need to engage in certification to stimulate demand, while firms offering the conventional product do not have incentives to commit fraud, as their claims would not be credible.

The third set of results relates the fines for fraud with firms' profits and social welfare.[2] We find that the presence of fraud (which arises only under intermediate fines) induces (weakly) lower equilibrium prices of the different varieties for firms than in settings with honest marketing. As consumers' trust is imperfect in the presence of (some) fraud, the equilibrium price for uncertified CSR is lower than in settings with perfect trust. This induces increased competition in the markets for conventional and certified CSR, reducing their respective prices. A related finding is that when the conventional product is available to consumers, the coexistence of fraud and certification induces a differentiation premium below the marginal cost of CSR, and a certification premium above the marginal certification cost. The former results from the strategic interaction between fraudulent and truthful firms, which reduces the expected willingness to pay for uncertified CSR. The latter is also derived from this uncertainty, providing an extra price premium for the certified version. Finally, we find that social welfare (weakly) increases when fines for fraud increase. This implies a first best scenario with large fines and no need for certification. A second best scenario is one with intermediate fines where (some) fraud may arise. We find that this scenario is socially preferred to an alternative one with low fines and no fraud where all CSR requires third-party validation. This result is interesting and may put the social

[2] Previous scholarship has neglected for long the welfare effects of CSR (Lyon and Maxwell, 2008), with some exceptions (Baksi and Bose, 2007; Bansal and Gangopadhyay, 2003; García-Gallego and Georgantzís, 2011; Moraga-González and Padrón-Fumero, 2002).

desirability of public subsidies for promoting massive adherence to certification programs into question.

In sum, our paper contributes to the literature on CSR by integrating self-differentiation investments and certification decisions in a single modeling framework, which in turn requires explicitly modeling the role of the potential for fraud on firms' strategies and consumers' decisions. The related literature up to date has studied these two features on self-differentiation and certification separately.

First, the literature focusing on incentives for self-differentiation (i.e., providing or not CSR) addresses the implications of production technology choices in markets where (at least a subset of) consumers value and can observe the CSR quality of products (e.g. Arora and Gangopadhyay, 1995; Conrad, 2005). Choosing technology acts as product positioning, opening up for differentiation strategies for firms (Kitzmueller and Shimshack, 2012). Substantial research efforts in these settings have addressed the interaction between the institutional context (taxes, subsidies and standards) with differentiation investments (Bansal and Gangopadhyay, 2003; Moraga-González and Padrón-Fumero, 2002). More generally, previous theoretical contributions claim that CSR markets can be understood as contexts where consumers can choose to consume impure public goods (in particular, for green products) that generate both private and public features (Kotchen, 2005, 2006, 2013). However, this literature has not considered the credence good nature of voluntary abatement efforts by firms.[3]

[3] Dulleck et al. (2011) characterize the credence good literature in two separate streams. The first considers that consumers do not know what they need, but they observe the utility from what they get. The subjects of interest in this case are markets, such as those for doctors and mechanics, where undertreatment, overtreatment and overcharging are crucial forms of fraud (e.g. Emons, 1997). Dulleck and Kerschbamer (2006) presents an excellent review of this literature jointly with a model of credence goods that provides an unifying framework for the different market institutions and information structures on efficiency. The second stream of the credence goods literature is based on the definition that credence goods have qualities that are difficult to judge even after purchase. In this case, consumers know what

Second, the literature focusing on certification (i.e., providing or not third party validation of CSR) addresses the asymmetric information problem characterizing the credence nature of some voluntary efforts by firms (e.g. Amacher et al., 2004; Sedjo and Swallow, 2002). A central aspect of this literature is that firms producing conventional products have incentives to pretend to be CSR (i.e. cheat) and profit from consumers' higher willingness to pay, while saving the associated production costs. An implication is that in the absence of credible information disclosure mechanisms, consumers will not believe firms' claims about the quality of products, requiring the use of certification. Available strategies for firms are therefore to invest in CSR and certify their product or use conventional (non-CSR) production. Also in this stream of literature, substantial research efforts have tackled the interrelation between voluntary certification and the institutional context (Baron, 2011; Feddersen and Gilligan, 2001; Glachant and Moineville, 2012; Heyes and Maxwell, 2004). Additionally, this literature has explored the implications of imperfect certification processes in which errors can occur (Mason, 2006) and in contexts where these errors are noisy (Mason, 2011). Then, as well, consumers' uncertainty about certification standards affect managerial decisions to have a product certified (Harbaugh et al., 2011). Baksi and Bose (2007) address the incentives for firms to *deliberately* misuse the labeling process to fraudulently sell their products as CSR. These authors compare the optimal monitoring policy by governments to deter fraud in equilibrium when the government imposes (forces non-voluntary) certification on low-quality firms, when it does so on high-quality firms, and when the low- and/or high-quality firms can self-differentiate their products. In Baksi and Bose, consumers' beliefs play no role, as monitoring by the government is intense enough to preclude the emergence of fraud in equilibrium. Glachant and Moineville (2012)

---

they want but observe neither what they get nor the utility derived from what they get. In this paper, we adhere to the second definition.

address the influence of consumers' beliefs on firms' claims for self-differentiation, but still in settings in which labeling is not voluntary. In Baksi and Bose (2007), the government forces certification of production practices, while in Glachant and Moineville (2012), NGO's news releases on firms behavior are the certification to consumers. In sum, none of these previous studies allows firms to choose between uncertified and certified CSR investments, as the present study does. This is relevant because as Kitzmueller and Shimshack (2012) argue, the notion of "voluntary behavior" is critical in the definition of CSR.[4]

We organize the remaining of the paper as follows. In the next section, we introduce the model. In Section 3, we present the characteristics of the equilibria configurations for different levels of the fines for fraud. In Section 4, we derive the implications of differences in fines on firms' production and marketing strategies, the emergence of fraud, equilibrium prices and social welfare. We conclude in Section 5. In the Appendix, we provide all the technical details.

2. The model

Suppose that a low (L) and a high (H) quality versions of a product are available in the market. We assume that each firm specializes in the production of either the low or the high quality version. All the consumers have the same willingness to pay for the low quality version, denoted as $\underline{\theta} > 0$, but they differ in their preferences for high quality. Let θ_j represent consumer j's additional willingness to pay for the high quality version, which is uniformly distributed in the interval $[0,1]$.

[4] Some examples of famous voluntary certification programs are the Global Reporting Initiative (GRI), Blue Angel, Forest Stewardship Council (FSC), ISO 14001, or the he EU Eco-Management and Audit Scheme (EMAS).

The product at hand presents credence good attributes and, therefore, it is impossible for consumers to distinguish between high and low quality even after purchase. Then, a firm producing the low quality version may be tempted to falsely market its product as high quality (L_H, where the subindex denotes the marketing strategy of the firm), aiming to capture additional consumer surplus, rather than truthfully market the product as low quality (L_L). Being aware of this, consumers form beliefs about the probability that an uncertified firm that claims to offer high quality actually does so, denoted as $\alpha \in [0,1]$. High quality can be guaranteed by a certification scheme[5] and consequently firms producing the high quality version can either market it as high quality (H_H) or certified high quality (H_C). Since firms producing the high quality version do not have any incentive to market the product as low quality, consumers perfectly identify that a product marketed as low quality is indeed low quality.

The product is supplied by a group of *N* firms. The number of firms following each strategy is represented by n_i, $i = L_L, L_H, H_H, H_C$, where $\sum_i n_i = N$. For a firm of type i, profits are defined as:

$$\pi_i = q_i(p_i - h_i - c_i - f_i),$$

where q_i is the total amount of the good produced and sold by a firm of type i, p_i is the per-unit price and (h_i, c_i, f_i) are, respectively, the per-unit cost of producing the high quality version, the per-unit certification cost and the per-unit expected fine for fraud. Without loss of generality, the per-unit cost of producing the low quality version is zero. For low quality firms marketing the product as low quality $h_{L_L} = c_{L_L} = f_{L_L} = 0$, for low quality firms falsely marketing the product as high quality $h_{L_H} = c_{L_H} = 0$, $f_{L_H} = f$, for

[5] We abstract from mistakes in the certification process addressed in Mason (2006; 2011) and consequently we assume that certification is a true signal of high quality.

high quality firms marketing the product as high quality $h_{H_H} = h$, $c_{H_H} = f_{H_H} = 0$, and for high quality firms marketing the product as certified high quality $h_{H_C} = h$, $c_{H_C} = c$, $f_{H_C} = 0$.

Therefore, there are four possible combinations of production and marketing strategies for the firms. However, consumers can only distinguish among the three marketed varieties, namely low, uncertified high, and certified high quality. In the process of deciding their production and marketing strategies, firms consider the consumers' surplus associated with each marketed variety. Consumers' surplus for the low quality version is simply the difference between the corresponding willingness to pay $\underline{\theta}$ and the price of the product p_{L_L}. Next, perfect credibility with regards to the certifying party ensures that consumers identify a certified high quality product as high quality. Then, consumer *j*'s surplus is the difference between the willingness to pay in this case, $\underline{\theta} + \theta_j$, and the price p_{H_C}. However, consumers cannot distinguish between fraudulent high quality (L_H) and truthful high quality (H_H), since both varieties are presented to the market as high quality and none is certified. In this case, consumers believe that the probability that the product is truly high quality is $\propto$. Assuming risk neutrality, consumer *j*'s expected surplus is the difference between the *expected* willingness to pay (i.e., $\propto \left(\overline{U} + \theta_j\right) + (1-\propto)\overline{U} = \overline{U} + \propto \theta_j$) and the corresponding price ($p_{L_H} = p_{H_H}$).

We assume that firms first decide the variety they produce, being a combination of the quality of the product and the marketing strategy, and then compete *á la Cournot*. The relevant equilibrium concept to solve this game with incomplete information is the perfect Bayesian equilibrium (see Fudenberg and Tirole, 1991). In this equilibrium, the strategies of the players are optimal for given beliefs and strategies of the other players.

In turn, beliefs on the equilibrium path are updated from observed actions according to Bayes' rule, while beliefs off the equilibrium path are updated according to Bayes' rule where possible.

3. Equilibria configurations

In this section, we characterize the different equilibria configurations, including the equilibrium prices, price premia, and the number of firms adopting each possible strategy. Two different price premia are relevant in this context, namely the *differentiation* and the *certification* price premia for the uncertified and certified high quality versions of the product, respectively. The former refers to the difference between the equilibrium prices of the uncertified high and the low quality versions, while the latter reflects the difference between certified and uncertified high quality prices. A complete characterization of all the possible types of equilibria can be found in the Appendix.

Three different scenarios that depend on the magnitude of the fines for fraud are relevant in studying the characteristics of the equilibria configurations. We refer to these three cases as of large, low and intermediate fines, respectively. We present the intuitive case of large fines in the first place.

3.1. Large fines for fraud

In this case, the per-unit expected fine for fraud is larger than the marginal cost of producing high quality ($f > h$). Proposition 1 presents the characteristics of a heterogeneous equilibrium where both the low and the high quality versions of the product can coexist in the market.

Proposition 1. *If $f > h$, both the low and the high quality versions of the product can coexist in equilibrium ($L_L H_H$). All firms producing low quality market the product as low quality (i.e., there is no fraud) and no firm offering high quality engages in certification. The respective equilibrium prices and number of firms pursuing each strategy are $p_{L_L} = \frac{1}{N}$, $p_{H_H} = \frac{1}{N} + h$, $n_{L_L} = hN$ and $n_{H_H} = (1-h)N$.*

Intuitively, if the expected fine is large enough to restrain firms producing the low quality version to commit fraud, all consumers (consistently) believe that the uncertified firms that claim to offer the high quality version do so truthfully (that is, consumers' belief in equilibrium is $\alpha^* = 1$). In this case, the (expected) willingness to pay for the uncertified high quality and the certified varieties coincide. Thus, certification has no influence in the demand, as it does not provide any additional information.

In terms of the number of firms following each strategy, the larger the marginal cost of producing high quality, the smaller the number of firms offering high quality. Consequently, the larger is the number of firms supplying low quality. In addition, the price of the low quality version is independent of the marginal cost of producing high quality (h), while the price of the high quality variety positively depends on such cost. Indeed, the differentiation price premium coincides exactly with the additional marginal cost of producing the high quality version (h).[6]

Finally, the likelihood of the equilibrium configuration $L_L H_H$ also depends on h being small enough. The explanation is as follows. The indifferent consumer between

[6] The equilibrium price of the low quality variety is independent of *h* due to two opposing effects that exactly compensate one for the other. On the one hand, the price of the low quality variety increases when *h* increases as a result of the interdependence with the price of the high quality version through the best response equation (see the Appendix for details). On the other hand, the number of firms offering the low quality version in equilibrium increases when *h* increases and therefore, competition in the market for the low quality variety increases, which reduces the corresponding price.

the low and the uncertified high quality varieties is simply $\theta^{LH} = \mathrm{p}_{\mathrm{H}_\mathrm{H}} - \mathrm{p}_{\mathrm{L}_\mathrm{L}} = h$.[7] Then, consumers with an additional willingness to pay for high quality lower than h purchase the low quality version, while consumers with an additional willingness to pay for high quality larger than h purchase the high quality version of the product. Thus, the heterogeneous equilibrium where the two versions of the product are offered exists as long as the indifferent consumer is within the feasible range, $\theta_j \in [0,1]$. Since $h > 0$, the critical condition in defining the existence of the heterogeneous equilibrium is $\theta^{LH} < 1$, which therefore implies $h < 1$. Intuitively, firms producing the low and the high quality versions of the product coexist in the market as long as there are consumers with an additional willingness to pay for high quality larger than the additional marginal production cost. Otherwise, all the firms supply the low quality version of the product, at the equilibrium price $p_{L_L} = \frac{1}{N}$.

3.2. Low fines for fraud

In this case, the per-unit expected fine for fraud is smaller than the marginal cost of producing the high quality version ($f < h$). Both the low and the high quality versions can coexist in the market as well, but now all the firms offering the high quality version must certify their product to signal high quality. The details of the heterogeneous equilibrium in this case are presented next.

Proposition 2. *If $f < h$, both the low and the high quality versions of the product can coexist in equilibrium (L_LH_C). All the firms producing low quality market the product as low quality (i.e., there is no fraud) and all the firms offering high quality certify their product. The respective equilibrium prices and number of firms pursuing*

[7] The corresponding consumer j's surpluses for the low and the high quality varieties are, respectively, $\underline{\theta} - p_{L_L}$ and $\underline{\theta} + \theta_j - p_{\mathrm{H}_\mathrm{H}}$.

each strategy are $p_{L_L} = \frac{1}{N}$, $p_{H_C} = \frac{1}{N} + h + c$, $n_{L_L} = (h + c)N$ *and* $n_{H_C} = (1 - h - c)N$.

Surprisingly, no firm offering the low quality version of the product engages in fraudulent marketing in equilibrium, even when the expected fine for fraud is very low. The reason is that consumers' beliefs are endogenous and updated in a Bayesian way. Firms producing the low quality version are obviously tempted to falsely market their product as high quality but they anticipate than by doing so all consumers will (consistently) believe that uncertified firms that claim to offer the high quality version are fraudulent. Therefore, the equilibrium consumers' belief is $\alpha^* = 0$ and uncertified claims cannot stimulate demand. In fact, under $\alpha^* = 0$, the expected willingness to pay for the low and the uncertified high quality varieties coincide. Consequently, the firms offering the high quality version need to certify their product to truthfully signal that they are producing high quality.[8]

The explanation of the characteristics of this equilibrium is analogous to the previous case under large fines. The only difference is that now, the price of the high quality version depends on $h + c$, that is, the sum of the marginal cost of producing high quality and the per-unit cost of certification. Then, the price premium for the certified high quality version, which includes both the differentiation and certification price premia, is exactly $h + c$. Since the mathematical expression for the indifferent consumer between the low and the certified high quality varieties is $\theta^{LC} = \mathrm{p_{H_C}} - \mathrm{p_{L_L}} = h + c$, , the heterogeneous equilibrium where the two versions of the product are offered

[8] Note that the belief $\propto^* = 0$ is off the equilibrium path, but anyway needed to sustain the equilibrium configuration where firms producing the low quality version never commit fraud and firms producing the high quality version always engage in certification.

exists as long as $h + c < 1$. Notice thus that with low fines the range of production costs for which firms produce high quality shrinks.

3.3. Intermediate fines for fraud

In this last scenario, the per-unit expected fine for fraud equals the marginal cost of producing the high quality version ($f = h$).[9] Again, the low and the high quality versions of the product can coexist as well, but now in some equilibria firms producing the low quality version commit fraud. The emergence of fraud depends on the marketing strategy selected by the firms producing the high quality version. The following proposition presents all the possible heterogeneous equilibrium configurations.

Proposition 3. *If $f = h$, both the low and the high quality versions of the product can coexist in equilibrium. Several equilibrium configurations arise in this case:*

(i) *$L_L H_C$: All the firms producing low quality market the product as low quality (i.e., there is no fraud), while all the firms offering high quality certify their product. The respective equilibrium prices and number of firms pursuing each strategy are $p_{L_L} = \frac{1}{N}$, $p_{H_C} = \frac{1}{N} + h + c$, $n_{L_L} = (h + c)N$ and $n_{H_C} = (1 - h - c)N$.*

(ii) *$L_L L_H H_H$: Some firms producing low quality fraudulently market their product as high quality, while no firm producing high quality engages in*

[9] We could alternatively assume that consumers are uncertain about the exact level of the expected fine for fraud (maybe because this fine depends on the probability of being caught, which might have a subjective component on the part of the firms), but they know that such an expected fine belongs to an interval $[\underline{f}, \bar{f}]$, $\underline{f} < \bar{f}$. Then, a scenario of large fines would be such that $h < \underline{f}$, intermediate fines would be such that $h \in [\underline{f}, \bar{f}]$ and, finally, low fines would be characterized by $h > \bar{f}$. The main results of the paper remain valid under this alternative specification.

certification. The respective equilibrium prices and number of firms pursuing each strategy are $p_{L_L} = \frac{\propto^*}{N}$, $p_{L_H} = p_{H_H} = \frac{\propto^*}{N} + h$, $n_{L_L} = \frac{hN}{\propto^*}$, $n_{L_H} = \frac{(1-\propto^*)(\propto^*-h)N}{\propto^*}$ *and* $n_{H_H} = (\propto^* - h)N$, *where* $\propto^* = (0,1)$.

(iii) $L_L L_H H_H H_C$*: Some firms producing low quality fraudulently market their product as high quality, while some firms producing high quality engage in certification. The respective equilibrium prices and number of firms pursuing each strategy are* $p_{L_L} = \frac{h\left[(1-\propto^*)^{1/2}+\propto^*-1\right]+c\left[\propto^*-(\propto^*)^{1/2}\right]+(\propto^*)^{1/2}(1-\propto^*)}{(1-\propto^*)^{1/2}\cdot\left\{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]\right\}}$, $p_{L_H} = p_{H_H} = (1-\propto^*)^{1/2} p_{L_L} + h$, $p_{H_C} = \left(\frac{1-\propto^*}{\propto^*}\right)^{1/2} p_{L_L} + h + c$, $n_{L_L} = \frac{h}{p_{L_L}} + (1-\propto^*)^{1/2} - 1$, $n_{L_H} = (1-\propto^*)\left\{\frac{\propto^*(h+c)-h}{p_{L_L}(1-\propto^*)^{1/2}} + (\propto^*)^{1/2} + (1-\propto^*)^{1/2} - 1\right\}$, $n_{H_H} = \propto^* \left\{\frac{\propto^*(h+c)-h}{p_{L_L}(1-\propto^*)^{1/2}} + (\propto^*)^{1/2} + (1-\propto^*)^{1/2} - 1\right\}$ *and* $n_{H_C} = \frac{(\propto^*)^{1/2}(1-\propto^*-c)}{p_{L_L}(1-\propto^*)^{1/2}} + (\propto^*)^{1/2} - 1$, *where* $\propto^* = (0,1)$.

(iv) $L_H H_H$*: All the firms producing low quality fraudulently market their product as high quality, while no firm producing high quality engages in certification. The respective equilibrium prices and number of firms pursuing each strategy are* $p_{L_H} = p_{H_H} = \frac{1-\propto^*}{N} + h$, $n_{L_H} = (1-\propto^*)N$ *and* $n_{H_H} = \propto^* N$, *where* $\propto^* = (0,1)$.

(v) $L_H H_H H_C$*: All the firms producing low quality fraudulently market their product as high quality, while some firms producing high quality engage in certification. The respective equilibrium prices and number of firms pursuing each strategy are* $p_{L_H} = p_{H_H} = \frac{1-\propto^*}{N} + h$, $p_{H_C} = \frac{1-\propto^*}{N} + h + c$, $n_{L_H} = cN$, $n_{H_H} = \frac{\propto^* cN}{1-\propto^*}$ *and* $n_{H_C} = \frac{(1-\propto^*-c)N}{1-\propto^*}$, *where* $\propto^* = (0,1)$.

In contrast with the two scenarios of large and low fines presented in Propositions 1 and 2 above, the present scenario is compatible with any consumers' belief in equilibrium, $\propto^* \in (0,1)$, and this opens up for the existence of all possible combinations of firms' strategies. In particular, the three marketing strategies (low, high and certified high quality) can coexist in this scenario (case iii of Proposition 3 characterizes this situation). Equilibrium $L_L H_C$ (case i) is the only one in which firms producing low quality do not commit any fraud. In this case, all the firms producing high quality engage in certification, which is incompatible with the presence of low quality firms committing fraud. Therefore the resulting equilibrium prices and number of firms following each strategy in case (i) are the same as those presented in Proposition 2. However, in cases (ii) – (v), some or even all the firms producing low quality fraudulently market their product as high quality. The key aspect that allows for this fact is the existence of firms producing high quality that do not certify their product.

Case (ii) in Proposition 3, equilibrium $L_L L_H H_H$ with no certification and with fraud from low quality firms, is similar in spirit to the one presented in Proposition 1, where no firm producing high quality engages in certification. The difference now is that, under risk neutrality, firms are indeed indifferent between producing the uncertified high quality version of the product or producing the low quality version and risking an expected penalty for fraud. Therefore, any consumers' belief is compatible in equilibrium, that is, $\propto^* \in (0,1)$. The equilibrium prices and number of firms following each strategy depend on the equilibrium consumers' belief. The lower the consumers' expectation $\propto^*$, the lower the prices of both the low and the (uncertified) high quality versions of the product. The reason is that consumers' additional (expected) willingness to pay for high quality is increasing in $\propto$, and the prices of the low and the high quality

varieties are positively correlated through the best response equation (see the Appendix for details).

Interestingly, the equilibrium $L_L L_H H_H H_C$ with high and low quality, fraud and certification (case iii of Proposition 3) presents idiosyncratic characteristics. In this case, the three marketing strategies (low, high and certified high quality) coexist in equilibrium. Also, some firms producing the low quality version find it optimal fraudulently market the product as high quality. As opposed to the cases previously analyzed, in this case the prices of *all* three varieties depend on the marginal cost of producing high quality and the marginal certification cost, since the prices of the three varieties interact through the respective best response equations.[10] Regarding the number of firms following the different strategies, the effect of h on n_{L_L}, n_{L_H} or n_{H_H} is uncertain, but the corresponding effect n_{H_C} is unambiguously negative for any $\propto^* \in (0,1)$. The higher the level of trust of consumers, the lower is the use of certification. The per-unit certification cost c also affects the equilibrium number of firms following each possible strategy. This effect is unambiguously positive on n_{L_L}, n_{L_H} and n_{H_H} for any $\propto^* \in (0,1)$ and, therefore, negative for n_{H_C}, as expected.

Next, in equilibrium $L_H H_H$ (case iv of Proposition 3) all the firms market their product as high quality without certification. Consumers do not observe any production quality differences, but all the firms that produce the low quality version falsely market the product as high quality. The number of firms following the fraudulent strategy

[10] Specifically, the prices of the low and the uncertified high quality varieties are increasing in the marginal cost of producing high quality (since $(1-\propto^*)^{1/2} + \propto^* - 1 > 0$ for any $\propto^* \in (0,1)$) and decreasing in the marginal certification cost (since $\propto^* - (\propto^*)^{1/2} < 0$ for any $\propto^* \in (0,1)$). On the contrary, the price of the certified high quality variety is increasing in both marginal costs (the latter is due to the fact that the term $\frac{\propto^* - (\propto^*)^{1/2}}{(1-\propto^*)^{1/2} \cdot \{N+3-2[(1-\propto^*)^{1/2} + (\propto^*)^{1/2}]\}}$ is negative but smaller than one in absolute value for any $\propto^* \in (0,1)$).

crucially depends on the consumers' equilibrium belief, $\propto^*$and, again, any $\propto^* \in (0,1)$ is compatible in equilibrium.

Finally, in equilibrium $L_H H_H H_C$ (case v of Proposition 3) all firms producing the low quality version commit fraud and market their product as high quality. In this case, uncertified and certified high quality coexist as the only marketing strategies. Interestingly, the consumers' equilibrium belief $\propto^*$negatively affects the price of both uncertified and certified high quality. Also, as expected, the prices of both varieties positively depend on the marginal cost of producing high quality, h, while the marginal certification cost only affects the certified high quality version of the product. Finally, this certification cost (and not the marginal cost of producing high quality) has a crucial effect on the number of firms following each strategy. The effect of the certification cost on n_{L_H} and n_{H_H} is positive, while the corresponding effect on n_{H_C} is negative.

Next, we analyze the conditions for the existence of each type of equilibrium configuration (see the Appendix for a detailed mathematical characterization of such conditions). Here, we illustrate the likelihood of each type of equilibrium by means of Figure 1. This figure represents, in the space of (h, c) combinations, the areas in which each possible type of equilibrium arises.[11, 12]

[INSERT FIGURE 1 ABOUT HERE]

[11] Since in this case we are assuming that $h = f$, every time we consider a change in h, an associated change of the same magnitude and sign is assumed with regards to f. This strategy allows us to analyze the effect of the value of h on the likelihood of each type of equilibrium configuration within the situation of intermediate fines. The comparison of the three ranges of the fines (which is possible only when we consider a change in h keeping f fixed) is postponed for the next section.

[12] The line $\theta^{HC} = 1$ defines the set of combinations (h, c) such that the indifferent consumer between uncertified and certified high quality has an additional willingness to pay for certified high quality of 1. Also, the line $\theta^{LH} = \theta^{HC}$ defines the set of combinations (h, c) such that the indifferent consumer between low and uncertified high quality and the indifferent consumer between uncertified and certified high quality coincide. Finally, the line $\theta^{LH} = 0$ defines the set of combinations (h, c) such that the indifferent consumer between low and uncertified high quality has an additional willingness to pay for high quality of zero.

As expected, in settings with large marginal costs of producing high quality and certification costs (i.e., large h and c), only firms producing the low quality version exist in equilibrium (area L_L in Figure 1). As h decreases (horizontal movement towards the vertical axis in Figure 1), it becomes possible for firms offering the low and high quality versions to coexist. Depending on the ratio of h to c, there can be low and certified high quality firms coexisting (case i, area $L_L H_C$) low and uncertified high quality firms coexisting (case ii, area$L_L L_H H_H$), or even the three marketing strategies coexisting (case iii, area $L_L L_H H_H H_C$). If h is small enough as compared with c (bottom left region of Figure 1), it is even possible that no firm markets its product as low quality, and only uncertified high quality is offered in the market (case iv, area $L_H H_H$) or uncertified and certified high quality with some fraud (case v, area $L_H H_H H_C$). The configuration of equilibria in Figure 1 is highly dependent on the number of firms in the industry (N). Figure 2 presents the likelihood of each possible equilibrium configuration for a large number of firms.

[INSERT FIGURE 2 ABOUT HERE]

By comparing Figures 1 and 2, the first thing we notice is that area L_L remains the same in both figures. This means that the likelihood of high quality production (being or not certified, and being or not truthful) is independent of the number of firms in the industry. Also, we can see that the line $\theta^{HC} = 1$ has pivoted down in Figure 2, while the line $\theta^{LH} = \theta^{HC}$ has pivoted up. In fact, it can be easily shown that as the number of competing firms increases, areas $L_H H_H$ and $L_H H_H H_C$ tend to disappear, since the line $\theta^{LH} = 0$ shifts to the left to eventually collapse to the vertical axis as N increases. This means that when the number of firms in the industry is large enough, there is always a subset of firms producing and marketing the low quality version of the

product. Note that the likelihood of fraud decreases when the number of firms in the industry increases (put differently, the sum of areas L_L and $L_L H_C$ is larger in Figure 2 than in Figure 1). Also, certified high quality is less likely when the size of the industry increases (the sum of areas $L_L H_C$, $L_H H_H H_C$ and $L_L L_H H_H H_C$ in Figure 1 is larger than the sum of areas $L_L H_C$ and $L_L L_H H_H H_C$ in Figure 2).

Finally, the likelihood of the corresponding equilibrium configurations also depends on the consumers' equilibrium belief, $\propto^*$. For example, in Figure 1[13], the area $L_L L_H H_H$ decreases if $\propto^*$ decreases ($h = \propto^*$ moves to the left and $c = 1 - \propto^*$ moves up) while the area $L_L H_C$ increases ($h = \propto^*$ moves to the left and $\theta^{LH} = \theta^{HC}$, or equivalentely $c = \frac{1-\propto^*}{\propto^*} h$, pivots to the left). Therefore, the set of combinations (h, c) under which certified high quality arises clearly increases when $\propto^*$ decreases. The likelihood of the coexistence of the three marketing strategies and fraud (or area $L_L L_H H_H H_C$) is maximized when $\propto^* = 1/2$.

4. How do (expeted) fines for fraud affect equilibrium outcomes?

In this section, we analyze the differences between the three ranges of the fines for fraud studied in the previous section with regards to the properties of the different types of equilibria. In particular, we analyze differences with regards to the likelihood of high quality production and certification, the likelihood of fraud, the resulting prices and price premia, and the social welfare levels.

Figure 3 compares the configuration of equilibria for the three cases of low, intermediate and large fines, respectively. The grey areas represent the presence of

[13] We refer to Figure 1 for illustration purposes, but the same kind of conclusions can be obtained using Figure 2.

certified high quality, while the dotted areas represent the presence of uncertified high quality. Uncertified and certified high quality coexist only in the case of intermediate fines, which is illustrated by the dotted-grey areas.

[INSERT FIGURE 3 ABOUT HERE]

It is easy to see that the likelihood of certified high quality (grey area) decreases as fines for fraud increase. Certification occurs for a broader combination of high quality production cost and certification costs in the case of low fines, shrinks for intermediate fines, to eventually disappear under large fines. This result has to do with the fact that, in equilibrium, consumers' beliefs on firms' claims are increasing in the level of the fines ($\propto^* = 0$ in the case of low fines, $\propto^* \in (0,1)$ in the intermediate case, and $\propto^* = 1$ in the case of large fines). As consumers' trust on uncertified claims increase, their additional willingness to pay for the certified version decreases. At the same time, higher consumers' trust on uncertified claims increases the presence of uncertified high quality (note that the dotted area is the largest under large fines, or complete trust, and it disappears under low fines, or nil trust). The reason is that consumers' additional willingness to pay for uncertified high quality, as compared to the low quality variety, increases as $\propto^*$ increases.

Combining these two effects, the production of high quality, being certified or not (and represented by the sum of the dark grey and dotted areas) is maximized in the case of large fines for fraud. Although certification is absent in this case, consumers' additional willingness to pay for the high quality version (relative to the low quality one) is the largest. Thus, firms have the largest incentive to produce high quality; not only consumers' additional willingness to pay is the largest but also production costs include only quality costs h and no certification costs. Conversely, in the case where

fines for fraud are low, the presence of certification practices is the largest, although this does not imply a larger likelihood of high quality production over all.[14]

All this analysis is summarized in the following:

Result 1. *As the fines for fraud increase, the range of parameter values for which there are firms supplying the high quality version of the product increases, at the expense of decreasing certification.*

We now analyze the link between the level of the fines and the presence of fraud in the industry. Figure 4 compares the three cases of low, intermediate and large fines. Shadowed areas represent equilibria configurations where fraud exists in equilibrium.

[INSERT FIGURE 4 ABOUT HERE]

Under large fines, firms have no incentives to commit fraud, since the expected fine for fraud is larger than the additional marginal cost of producing high quality. This induces consumers to completely trust firms' claims ($\propto^* = 1$). In the opposite case of low fines, firms do not commit fraud either, but the explanation is completely different. In this situation, all the firms producing the low quality version are tempted to commit fraud, since the expected fine for fraud is lower than the additional marginal cost of producing high quality. But consumers anticipate these incentives and will therefore form the belief that no firm offering uncertified high quality is to be trusted ($\propto^* = 0$). As a result, firms producing high quality need to engage in certification to credibly signal high quality.

[14] The reader can easily check that all this analysis remains valid when the number of firms in the industry is large enough. Figures 3a and 3c are independent of N, while Figure 3b has to be appropriately updated according to Figure 2.

Fraud is then only possible in the intermediate case, where consumers' equilibrium level of trust on uncertified varieties can be anywhere between 0 and 1. For intermediate fines, fraud is present in all equilibrium configurations except in the areas L_L and $L_L H_C$. In the shadowed area of Figure 4, the marginal cost of producing high quality is large enough relative to the marginal cost of certification as for fraud to emerge. These results regarding fraud are then summarized next.

Result 2*. Fraud can only arise under intermediate fines. The key features that make fraud likely are the existence of firms producing uncertified high quality in equilibrium and less than complete consumers' trust on the quality claims of firms.*

Next, we analyze the impact of the level of fines for fraud on the equilibrium prices. The interesting cases to study are the combinations of (*h,c*) that result in different equilibria depending on the level of the fines. These areas can be identified by comparing the different panels in figures 3 and 4.[15]

The equilibrium price for the uncertified high quality version of the product is (weakly) lower with intermediate fines than with either of the other two extremes fines for fraud. The scenario of intermediate fines is the only one where fraud is possible and consequently consumers' level of trust for uncertified high quality is not complete. This results in a lower willingness to pay for uncertified high quality as compared to the situation of large fines. This in turn induces a decrease in the number of firms offering uncertified high quality and the corresponding increase in the number of firms offering either the low or the certified high quality versions. The result is increased competition

[15] We find five relevant combinations of equilibrium configurations, depending on the parameters (*h,c*). Each configuration combination is characterized by the configuration of the low, intermediate and large fines, respectively. The relevant combinations are: (i) $L_L H_C - L_L H_C - L_L H_H$; (ii) $L_L H_C - L_L L_H H_H - L_L H_H$; (iii) $L_L H_C - L_L L_H H_H H_C - L_L H_H$; (iv) $L_L H_C - L_L H_H - L_L H_H$; and, finally, (v) $L_L H_C - L_H H_H H_C - L_L H_H$. The two last cases are not possible when N is large enough.

in both markets, reducing the respective prices. Result 3 now summarizes this finding. (A complete proof of this result is offered in the Appendix.)

Result 3. *The equilibrium prices of the different varieties of the product in the case of intermediate fines are weakly lower than the corresponding prices of the different marketed varieties under large and low fines.*

Now, we compare the respective price premia in the three scenarios of low, intermediate and large fines. A direct comparison of Propositions 1, 2 and 3 shows that the respective differentiation and certification premia are in most settings the additional marginal cost of producing high quality (h) and the per-unit certification cost (c), respectively. A remarkable exception is equilibrium $L_LL_HH_HH_C$ for intermediate fines. There, the differentiation premium ($p_{H_H} - p_{L_L}$) is below the marginal cost of producing high quality (h), while the certification premium ($p_{H_C} - p_{H_H}$) is above the marginal certification fee (c).[16] This result illustrates that the strategic interaction between fraudulent and trustworthy products marketed as high quality reduces the expected willingness to pay for uncertified high quality, due to a reduction of consumers' trust on uncertified products. This uncertainty on firms' quality claims results in an extraordinary price premium for certified products beyond their certification costs. This result is summarized next.

Result 4. *When the low quality version is available to consumers, the coexistence of fraud and certification induces a differentiation premium below the*

16 The first result is due to the fact that $p_{H_H} - p_{L_L} = \left[(1-\alpha^*)^{1/2} - 1\right]p_{L_L} + h$ and $\alpha^* < 1$, which implies $\left[(1-\alpha^*)^{1/2} - 1\right] < 0$ and $p_{H_H} - p_{L_L} < h$. Regarding the second result, $p_{H_C} - p_{H_H} = (1-\alpha^*)^{1/2}\left[\frac{1}{(\alpha^*)^{1/2}} - 1\right]p_{L_L} + c$ and $\alpha^* < 1$, which implies $(1-\alpha^*)^{1/2}\left[\frac{1}{(\alpha^*)^{1/2}} - 1\right] > 0$ and $p_{H_C} - p_{H_H} > c$. Clearly, these price premia converge to the respective additional marginal costs, h and c, when the number of firms in the industry is large enough.

marginal production cost of high quality, and a certification premium above the marginal certification cost.

Finally, we compare the social welfare levels obtained under the three different ranges for the fines for fraud. In doing so, we consider the sum of consumers' surpluses and firms' profits. Since producers collect the prices paid by consumers, social welfare is the sum of the differences between consumers' willingness to pay and per-unit production (and certification, if applicable) costs.

The only possible heterogeneous equilibrium configuration under large fines is $L_L H_H$ (see Proposition 1 and Figure 3c). In this equilibrium, consumers with additional willingness to pay for high quality between 0 and h purchase the low quality version, while the remaining consumers (those with additional willingness to pay for high quality between h and 1) purchase high quality. The total cost of high quality in this case is h, since high quality is uncertified. Then, social welfare in this equilibrium can be written as follows:

$$\mathrm{SW}(\mathrm{L_L H_H}) = \int_0^{\mathrm{h}} \underline{\theta}\, \mathrm{d}\theta_{\mathrm{j}} + \int_{\mathrm{h}}^{1} (\underline{\theta} + \theta_{\mathrm{j}} - \mathrm{h})\mathrm{d}\theta_{\mathrm{j}} = \underline{\theta} + \frac{(1-\mathrm{h})^2}{2}$$

In the case of low fines (Proposition 2 and Figure 3a), the only possible equilibrium is $L_L H_C$. In this equilibrium, consumers with additional willingness to pay for high quality between 0 and $h + c$ purchase the low quality version, while the remaining consumers (those with additional willingness to pay for high quality between $h + c$ and 1) purchase high quality. The cost of high quality in this case is $h + c$, since high quality must be certified. Then, social welfare in this equilibrium is:

$$\mathrm{SW}(\mathrm{L_L H_C}) = \int_0^{\mathrm{h+c}} \underline{\theta}\, \mathrm{d}\theta_{\mathrm{j}} + \int_{\mathrm{h+c}}^{1} (\underline{\theta} + \theta_{\mathrm{j}} - \mathrm{h} - \mathrm{c})\mathrm{d}\theta_{\mathrm{j}} = \underline{\theta} + \frac{(1-\mathrm{h}-\mathrm{c})^2}{2}$$

A simple comparison of the social welfare evaluations under large and low fines lead us to conclude that social welfare is larger in the first case. The reason is very simple: high quality production is more expensive under low fines since firms must incur in certification costs to outstand consumers' trust.

Now we analyze social welfare in the case of intermediate fines (Proposition 3 and Figure 2). For ease of exposition, we consider a sufficiently large number of firms in the industry.[17] In that case, equilibria $L_L H_C$, $L_L L_H H_H$ and $L_L L_H H_H H_C$ are possible. First, the social evaluation of equilibrium $L_L H_C$ under intermediate fines is identical to that under small fines, and therefore, smaller than under large fines.

Second, in equilibrium $L_L L_H H_H$, consumers with additional willingness to pay for high quality between 0 and $\frac{h}{\propto^*}$ purchase the low quality version, while the remaining consumers (those with additional willingness to pay for high quality between $\frac{h}{\propto^*}$ and 1) purchase the product labeled as high quality. The cost of high quality in this case is h, since high quality is uncertified, but only a proportion $\propto^*$ of all producers marketing the product as high quality truly incur the cost. Then, (expected) social welfare in this case is:

$$\mathrm{SW}(\mathrm{L_L L_H H_H}) = \int_0^{\mathrm{h}/\propto^*} \underline{\theta}\, \mathrm{d}\theta_\mathrm{j} + \int_{\mathrm{h}/\propto^*}^1 (\underline{\theta} + \propto^* \theta_\mathrm{j} - \propto^* \mathrm{h})\mathrm{d}\theta_\mathrm{j} = \underline{\theta} + \propto^* \frac{(1 - \mathrm{h}/\propto^*)^2}{2}$$

Then, a comparison of $\mathrm{SW}(\mathrm{L_L L_H H_H})$ with the social welfare associated to the corresponding equilibrium under large fines, $\mathrm{SW}(\mathrm{L_L H_H})$, shows that $\mathrm{SW}(\mathrm{L_L H_H}) > \mathrm{SW}(\mathrm{L_L L_H H_H})$ since $\propto^* < 1$. Thus, social welfare is higher for large fines than for intermediate fines. In addition, a comparison of $\mathrm{SW}(\mathrm{L_L L_H H_H})$ with the social welfare

[17] The same results hold when the number of firms is small, but they are more tedious to present, since additional equilibria (iv) and (v) characterized in Proposition 3 can exist as well.

associated to the corresponding equilibrium for low fines, $SW(L_L H_C)$, shows that $SW(L_L L_H H_H) > SW(L_L H_C)$.[18] In sum, for the range of parameter values for which this equilibrium exists, the social welfare associated to intermediate fines (where there is fraud) lies between the respective evaluations under low and large fines.

Finally, in equilibrium $L_L L_H H_H H_C$, consumers with additional willingness to pay for high quality between 0 and $\frac{h}{\propto^*}$ purchase the low quality version, consumers with additional willingness to pay for high quality between $\frac{h}{\propto^*}$ and $\frac{c}{1-\propto^*}$ purchase uncertified high quality, and the remaining consumers (those with additional willingness to pay for high quality between $\frac{c}{1-\propto^*}$ and 1 purchase the certified high quality version). The cost of high quality in this case is h, since high quality is uncertified, but only a proportion $\propto^*$ of all producers marketing the product as high quality truly incur the cost. Then, the social welfare evaluation in this case is the following:

$$SW(L_L L_H H_H H_C)$$

$$= \int_0^{\frac{h}{\propto^*}} \underline{\theta}\, d\theta_j + \int_{\frac{h}{\propto^*}}^{\frac{c}{1-\propto^*}} (\underline{\theta} + \propto^* \theta_j - \propto^* h) d\theta_j + \int_{\frac{c}{1-\propto^*}}^{1} (\underline{\theta} + \theta_j - h - c) d\theta_j$$

$$= \underline{\theta} + \frac{1}{2} + \frac{c^2}{2(1-\propto^*)} + \frac{h^2}{2\propto^*} - h - c$$

Now, comparing social welfare under large and intermediate fines, it is the case that $SW(L_L H_H) > SW(L_L L_H H_H H_C)$, since $c < 1 - \alpha^*$. Also, comparing social welfare under intermediate and low fines, $SW(L_L L_H H_H H_C) > SW(L_L H_C)$, since $c > \frac{(1-\alpha^*)h}{\alpha^*}$.

[18] The reason is that $\propto^* \frac{(1-h/\propto^*)^2}{2} > \frac{(1-h-c)^2}{2}$ is equivalent to $(\propto^* - h) > (\propto^*)^{1/2}(1-h-c)$, which implies $c > \frac{(h+(\propto^*)^{1/2})(1-(\propto^*)^{1/2})}{(\propto^*)^{1/2}}$. This condition holds for any (h, c) for which the equilibrium $L_L L_H H_H$ is defined, i.e., $h < \propto^*$ and $c > 1 - \propto^*$.

Thus, the social welfare evaluation for intermediate fines also lies in this case between the respective evaluations under low and large fines.

Thus, the comparison of the respective social welfare levels under the three possible scenarios for the fines unveils the following result:

Result 5. *Ceteris paribus, social welfare levels (weakly) increase with the level of the fines for fraud.*

The implication of this result is that social welfare is maximized under large fines and minimized under low fines. In particular, the scenario of intermediate fraud fines, the only one where fraud can arise, results in social welfare levels in between the two extreme scenarios for the level of the fines. Thus, for large fines for fraud, not only the industry is more strongly investing in CSR efforts, resulting in the positive externalities associated to the credence good, but also social welfare is the largest. It is important to highlight that this is the scenario in which firms dispense with certification. At the minimum, result 5 questions the social desirability of policies that promote third-party certification as compared to policies that fight the prevalence of fraud.[19]

5. Conclusions

The coexistence of conventional production, CSR practices, fraud and certification is common in many market settings. In this paper, we have presented an oligopoly model of vertical product differentiation that formalizes CSR in settings where firms can

[19] This analysis abstracts from the potential disutility that consumers might associate to the existence of fraud in the industry. Yet, given that both for the scenarios of large and low fines for fraud no firm engages in fraudulent marketing, it would still be the case that social welfare would be maximized under large fines for fraud. Considering a disutility associated to the existence of fraud would potentially change only the ranking of social desirability between scenarios of intermediate or low fines for fraud.

choose to certify their responsible behavior to outstand in credibility with respect to other firms that truthfully (or not) claim to be socially responsible. We have treated CSR as a high quality version of a product, while conventional production has represented the low quality version. Thus, the model is equally applicable to a broad range of CSR practices, such as employment discrimination, "fair trade", respect for human rights, use of pesticides, genetically modified organism, and other contexts where high quality presents attributes of a credence good.

The model presents a broader strategy set for firms than those previously discussed in the literature in an integrated framework of analysis. First, firms can market themselves as responsible, even if they are not. This opens up for fraudulent marketing by firms, falsely claiming to be socially responsible. Second, firms' CSR strategies can include, but are not restricted to, certification. Thus, firms can choose to market their products as conventional, self-reported CSR, or certified CSR. And third, consumers hold endogenous beliefs on the prevalence of fraud in the industry and in equilibrium these expectations are correct.

Our results show that the interaction between the possible presence of fraud and certification critically alters market structure. We have found that increasing fines for fraud extends the likelihood of firms supplying CSR production, at the expense of a decreased likelihood of certification. Interestingly, the potential for fraud does not prevent the co-existence of fraudulent and honest CSR efforts. In fact, as fraud regulation and control increases (inducing in turn higher consumers' trust on firms' self-reported claims), the range of parameter values for which firms voluntary invest in CSR increases and so does the equilibrium number of firms that engage in CSR. Certification becomes a less likely strategy, whereas producing uncertified CSR is a more likely strategy, over-compensating the reduced likelihood of certified efforts. Thus, an

interesting implication of this first result is that it is not necessarily the case that widespread certification is synonymous of a more responsible industry.

A relevant finding of our study is that fraud only arises in settings with intermediate fines for fraud and where conventional and uncertified CSR production coexist in the market. The key aspect for the emergence of fraud is consumers' imperfect trust on the uncertified variety. Interestingly though, with low fines for fraud consumers can anticipate that all self-reports on CSR practices that are not third-party certified will be fraudulent. Thus, fraud is deterred through low trust on these claims.

Importantly, the existence of fraud comes at a cost for producers. Fraud induces (weakly) lower equilibrium prices of the different varieties than those when there is no fraud. Also, the coexistence of fraud and certification induces a differentiation premium below the marginal production cost of CSR, and a certification premium above the marginal certification cost. The former results from the strategic interaction between fraudulent and truthful firms that reduces the expected willingness to pay for uncertified CSR. As a result, firms that incur in fraud have higher actual profit margins than the firms that truthfully reveal their CSR investments but do not certify so. The latter also derives from consumers' uncertainty, providing an extra price premium for the certified version. Lastly, our results also shed light on the welfare implications of fraud. Social welfare (weakly) increases with increasing fines for fraud.

In sum, our findings support that policies fighting fraudulent CSR marketing simultaneously increase social welfare and guarantee the highest social responsibility of the industry supported by voluntary action. Yet, in many instances, governments behave *as if* adherence to certification programs was a socially desirable objective, as shown by the provision of direct and indirect subsidies for firms to become more responsible and

join certifications. In light of our results, markets where certification is widespread are not necessarily those where production is more responsible, because this is mainly needed in contexts where consumers do not trust firms' claims (i.e., contexts of low or intermediate fines for fraud). The need to invest in the extra-costs of certification, as compared to only investing in socially enhancing responsible production, compromises the overall social welfare outcomes.

It is worth noting that we have conducted our analysis assuming that fraud can only be associated to firms' uncertified CSR claims. Thus, we have assumed throughout that certification eliminates the asymmetric information problem of consumers, providing a true signal of a product quality. An interesting area for further research would be to allow for fraud also in the use of the certification. This would require a more complex modeling of consumers' beliefs.

References

Amacher, G.S., Koskela, E., Ollikainen, M., 2004. Environmental quality competition and eco-labeling. Journal of Environmental Economics and Management, 47, 284-306.

Arora, S., Gangopadhyay, S., 1995. Toward a theoretical model of voluntary overcompliance. Journal of Economic Behavior and Organization, 28, 289-309.

Baksi, S., Bose, P., 2007. Credence Goods, Efficient Labelling Policies, and Regulatory Enforcement. Environmental and Resource Economics, 37, 411-430.

Bansal, S., Gangopadhyay, S., 2003. Tax/subsidy policies in the presence of environmentally aware consumers. Journal of Environmental Economics and Management, 45, 333-355.

Baron, D.P., 2011. Credence attributes, voluntary organizations, and social pressure. Journal of Public Economics, 95, 1331-1338.

Commission, E., 2009, Europeans' attitudes towards the issue of sustainable consumption and production. Flash Eurobarometer Series #256. European Commission.

Conrad, K., 2005. Price Competition and Product Differentiation When Consumers Care for the Environment. Environmental and Resource Economics, 31, 1-19.

Dulleck, U., Kerschbamer, R., 2006. On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods. Journal of Economic Literature, 44, 5-42.

Dulleck, U., Kerschbamer, R., Sutter, M., 2011. The Economics of Credence Goods: An Experiment on the Role of Liability, Verifiability, Reputation, and Competition. American Economic Review, 101, 526-555.

Emons, W., 1997. Credence goods and fraudulent experts. The RAND Journal of Economics, 28, 107-119.

Feddersen, T.J., Gilligan, T.W., 2001. Saints and Markets: Activists and the Supply of Credence Goods. Journal of Economics & Management Strategy, 10, 149-171.

Fudenberg, D., Tirole, J., 1991. Game Theory. Cambridge: MIT Press.

García-Gallego, A., Georgantzís, N., 2011. Good and Bad Increases in Ecological Awareness: Environmental Differentiation Revisited. Strategic Behavior and the Environment:, 1, 7-88.

Glachant, M., Moineville, G., 2012, The informational role of non-governmental organization to induce self-regulation: Cheering the leaders or booing the laggards? Working paper series CERNA, MINES ParisTech., pp. 25.

Harbaugh, R., Maxwell, J.W., Roussillon, B., 2011. Label Confusion: The Groucho Effect of Uncertain Standards. Management Science, 57, 1512-1527.

Heyes, A.G., Maxwell, J.W., 2004. Private vs. public regulation: political economy of the international environment. Journal of Environmental Economics and Management, 48, 978-996.

Khanna, M., 2001. Non-Mandatory Approaches to Environmental Protection. Journal of Economic Surveys, 15, 291-324.

Kitzmueller, M., Shimshack, J., 2012. Economic Perspectives on Corporate Social Responsibility. Journal of Economic Literature, 50, 51-84.

Kotchen, M.J., 2005. Impure public goods and the comparative statics of environmentally friendly consumption. Journal of Environmental Economics and Management, 49, 281-300.

Kotchen, M.J., 2006. Green Markets and Private Provision of Public Goods. Journal of Political Economy, 114, 816-834.

Kotchen, M.J., 2013. Voluntary- and Information-Based Approaches to Environmental Management: A Public Economics Perspective. Review of Environmental Economics and Policy, 7, 276-295.

Lyon, T.P., Maxwell, J.W., 2008. Corporate Social Responsibility and the Environment: A Theoretical Perspective. Review of Environmental Economics and Policy, 2, 240-260.

Lyon, T.P., Maxwell, J.W., 2011. Greenwash: Corporate Environmental Disclosure under Threat of Audit. Journal of Economics & Management Strategy, 20, 3-41.

Mason, C., 2006. An economic model of ecolabeling. Environmental Modeling and Assessment, 11, 131-143.

Mason, C.F., 2011. Eco-Labeling and Market Equilibria with Noisy Certification Tests. Environmental and Resource Economics, 48, 537-560.

Moraga-González, J.L., Padrón-Fumero, N., 2002. Environmental Policy in a Green Market. Environmental and Resource Economics, 22, 419-447.

Portney, P.R., 2008. The (Not So) New Corporate Social Responsibility: An Empirical Perspective. Review of Environmental Economics and Policy, 2, 261-275.

Sedjo, R.A., Swallow, S.K., 2002. Voluntary Eco-Labeling and the Price Premium. Land Economics, 78, 272-284.

Mathematical Appendix

Proof of Proposition 1. If $f > h$, no firm producing the low quality version has incentives to falsely market the product as high quality. Consistently, consumers believe that the uncertified firms that claim to offer the high quality version do so truthfully, i.e., $\alpha^* = 1$.

Consumer *j*'s surplus associated with the low, high and certified high quality versions of the product when $\propto^* = 1$ are respectively $\underline{\theta} - p_{L_L}$; $\underline{\theta} + \theta_j - p_{H_H}$; $\underline{\theta} + \theta_j - p_{H_C}$. First note that the (expected) willingness to pay for the uncertified and certified high quality varieties coincide. As a result, no firm offering the high quality version has incentives to certify the product in equilibrium. Thus, only the low and uncertified high quality versions of the product are supplied, and the expression for the indifferent consumer between buying any of the two varieties is $\theta_j = p_{H_H} - p_{L_L}$. Therefore, $Q_L = p_{H_H} - p_{L_L;}$ is the aggregate demand for the low quality variety while $Q_H = 1 - p_{H_H} + p_{L_L}$ is the demand for the uncertified high quality version of the product.

Firms first decide their production and marketing strategies (in this case, either producing low quality and marketing it as low quality, or producing high quality and marketing it as uncertified high quality) and then they compete in quantities. Starting with the last stage, a low quality firm (denoted with subscript bi) selects the quantity so as to maximize individual profits as follows:

$\max_{q_{L_L}} \pi_{L_L} = q_{L_L} p_{L_L}$, where $p_{L_L} = p_{H_H} - Q_L$.

Since all firms offering the low quality version are equal, the first order condition of the maximization problem can be written as $q_{L_L} = \frac{p_{H_H}}{n_{L_L}+1}$. Substituting this into the demand function leads to the best response price $p_{L_L} = \frac{p_{H_H}}{n_{L_L}+1}$.

Similarly, a firm offering the uncertified high quality variety selects the quantity supplied that maximize individual profits:

$\max_{q_{H_H}} \pi_{H_H} = q_{H_H}(p_{H_H} - h)$, where $p_{H_H} = 1 + p_{L_L} - Q_H$.

Now, the first order condition of the maximization problem can be written as $q_{H_H} = \frac{1+p_{L_L}-h}{n_{H_H}+1}$. Substituting this into the corresponding aggregate demand function results the best response price $p_{H_H} = \frac{1+hn_{H_H}+p_{L_L}}{n_{H_H}+1}$.

As a result, the unique equilibrium is the solution of the two (best response) equations, $p_{L_L} = \frac{p_{L_L}}{n_{L_L}+1}$ and $p_{H_H} = \frac{1+hn_{H_H}+p_{L_L}}{n_{H_H}+1}$. This results in equilibrium prices conditional on the number of firms following each strategy, as follows:

$$p_{L_L}|(n_{L_L}, n_{H_H}) = \frac{1+hn_{H_H}}{(n_{H_H}+1)(n_{L_L}+1)-1}\ ;\ p_{H_H}|(n_{L_L}, n_{H_H}) = (n_{L_L}+1)\frac{1+hn_{H_H}}{(n_{H_H}+1)(n_{L_L}+1)-1}.$$

Now, to obtain the number of firms in equilibrium following each strategy we impose the condition that in equilibrium $\pi_{L_L} = \pi_{H_H}$, so that no firm has incentives to deviate to the other possible strategy. Substituting the equilibrium prices conditional on quantities in the profit function for low quality firms, $\pi_{L_L} = q_{L_L}p_{L_L}$, results in the expression $\pi_{L_L} = \frac{[1+hn_{H_H}]^2}{[(n_{H_H}+1)(n_{L_L}+1)-1]^2}$. Similarly, the profit function of a firm offering the uncertified high quality variety results in $\pi_{H_H} = \frac{[n_{L_L}+1-hn_{L_L}]^2}{[(n_{H_H}+1)(n_{L_L}+1)-1]^2}$. Therefore, the

equilibrium condition $\pi_{L_L} = \pi_{H_H}$ holds as long as $1 + hn_{H_H} = n_{L_L} + 1 - hn_{L_L}$, which results in equilibrium numbers of firms $n_{L_L} = hN$; $n_{H_H} = (1-h)N$. Substituting these results into the equilibrium prices conditional on the number of firms obtained above, we obtain the equilibrium prices $p_{L_L} = \frac{1}{N}$; $p_{H_H} = \frac{1}{N} + h$.

Proof of Proposition 2. Since $f < h$, consumers' equilibrium belief is 0, and only low and certified high quality firms coexist in the market. The cost of offering high quality is now $h + c$ instead of h. The rest of the proof is analogous to that of Proposition 1 and, therefore, omitted.

Proof of Proposition 3. If $h = f$, any $\propto^* \in (0,1)$ can be an equilibrium belief, and this allows for all the possible types of heterogeneous equilibrium configurations to exist.

(i). Only low and certified high quality are offered in equilibrium. The equilibrium belief $\propto^* \in (0,1)$ is off the equilibrium path and the result is the same as that found under Proposition 2.

(ii). Only low and uncertified high quality are offered in equilibrium. In this case, given consumers' equilibrium belief $\propto^* \in (0,1)$, the same proportion $\propto^*$ of firms labeling the product as of high quality are truly producing high quality. Since consumers cannot distinguish between true and fake high quality, the equilibrium price in both cases is the same, i.e., $p_{L_H} = p_{H_H}$. Following a similar procedure to that of Proposition 1, we first obtain the corresponding best response prices resulting from the interaction of low and uncertified high quality firms. Now, the demand for the low quality variety is $Q_L = \frac{p_{H_H} - p_{L_L}}{\propto^*}$, and the best response price of the low quality firms is $p_{L_L} = \frac{p_{H_H}}{n_{L_L}+1}$. Then, the demand for uncertified high quality is now $Q_H = 1 - \frac{p_{H_H} - p_{L_L}}{\propto^*}$,

which results in the best response price $p_{H_H} = p_{L_H} = \frac{\propto^* + p_{L_L} + h n_{H_H}}{n_{H_H}+1}$. Combining both reaction functions, we obtain $p_{L_L} = \frac{\propto^* + h\,(n_{L_H}+n_{H_H})}{(n_{L_L}+1)(n_{L_H}+n_{H_H}+1)-1}$; $p_{H_H} = p_{L_H} = \frac{(n_{L_L}+1)(\propto^* + h\,(n_{L_H}+n_{H_H}))}{(n_{L_L}+1)(n_{L_H}+n_{H_H}+1)-1}$, which results in the corresponding profit functions $\pi_{L_L} = \frac{[\propto^* + h\,(n_{L_H}+n_{H_H})]^2}{\propto^*[(n_{L_L}+1)(n_{L_H}+n_{H_H}+1)-1]^2}$; $\pi_{L_H} = \pi_{H_H} = \frac{[(n_{L_L}+1)\propto^* - h n_{L_L}]^2}{\propto^*[(n_{L_L}+1)(n_{L_H}+n_{H_H}+1)-1]^2}$.

Imposing the equilibrium condition $\pi_{L_L} = \pi_{L_H} = \pi_{H_H}$ and since $n_{L_L} + n_{L_H} + n_{H_H} = N$ and $n_{H_H} = \propto^* (\mathrm{n_{L_H}} + \mathrm{n_{H_H}})$, we obtain that the numbers of firms following the respective strategies in equilibrium are, respectively, $\mathrm{n_{L_L}} = \frac{\mathrm{hN}}{\propto^*}$, $\mathrm{n_{L_H}} = \frac{(1-\propto^*)(\propto^* - \mathrm{h})\mathrm{N}}{\propto^*}$ and $\mathrm{n_{H_H}} = (\propto^* - \mathrm{h})\mathrm{N}$, which result in the equilibrium prices $p_{L_L} = \frac{\propto^*}{N}$, $p_{L_H} = p_{H_H} = \frac{\propto^*}{N} + h$.

(iii). Low, uncertified and certified high quality are offered in equilibrium. In this case, given consumers' equilibrium belief $\propto^* \in (0,1)$, the same proportion $\propto^*$ of firms labeling the product as of uncertified high quality are truly producing high quality. Since consumers cannot distinguish between true and fake uncertified high quality, the equilibrium price in both cases is the same, i.e., $p_{L_H} = p_{H_H}$. Now, the corresponding expressions for consumers' surpluses with the low, high and certified high quality versions of the product for $\propto^* \in (0,1)$ are respectively $\underline{\theta} - p_{L_L}$; $\underline{\theta} + \alpha^* \theta_j - p_{H_H}$; $\underline{\theta} + \theta_j - p_{H_C}$. Therefore, the aggregate demands for the respective varieties of the product are the following:

$$Q_L = \frac{p_{H_H} - p_{L_L}}{\propto^*};\ Q_H = \frac{p_{H_C} - p_{H_H}}{1-\propto^*} - \frac{p_{H_H} - p_{L_L}}{\propto^*};\ Q_C = 1 - \frac{p_{H_C} - p_{H_H}}{1-\propto^*}.$$

Firms first decide their environmental strategy and then they compete in quantities. Starting with the last stage, a firm producing and marketing the product as low quality selects the quantity so as to maximize profits, as follows:

$\max_{q_{L_L}} \pi_{L_L} = q_{L_L} p_{L_L}$, where $p_{L_L} = p_{H_H} - \propto^* Q_L$.

Since all firms are equal, the first order condition of the maximization problem can be written as $q_{L_L} = \frac{p_{H_H}}{\propto^* (n_{L_L}+1)}$. Substituting this into the demand function leads to the best response price $p_{L_L} = \frac{p_{H_H}}{n_{L_L}+1}$. Thus, $q_{L_L} = \frac{p_{L_L}}{\propto^*}$, which results in $\pi_{L_L} = \frac{p_{L_L}^2}{\propto^*}$.

Similarly, honest (and also fake) firms offering uncertified high quality select the quantity that maximizes individual profits as follows:

$\max_{q_{H_H}} \pi_{H_H} = q_{H_H}(p_{H_H} - h)$, $(\max_{q_{L_H}} \pi_{L_H} = q_{L_H}(p_{L_H} - f))$

where $p_{H_H} = p_{L_H} = \propto^* p_{H_C} + (1 - \propto^*) p_{L_L} - \propto^* (1 - \propto^*) Q_H$

Since all firms selecting each strategy are equal, $p_{H_H} = p_{L_H}$ and $h = f$, the first order condition of the maximization problem can be written as $q_{L_H} = q_{H_H} = \frac{\propto^* p_{H_C} + (1-\propto^*) p_{L_L} - h}{\propto^* (1-\propto^*)(n_{L_H}+n_{H_H}+1)}$. Substituting this into the corresponding aggregate demand function leads to the best response price $p_{H_H} = p_{L_H} = \frac{\propto^* p_{H_C} + (1-\propto^*) p_{L_L} + h(n_{L_H}+n_{H_H})}{n_{L_H}+n_{H_H}+1}$. Then, $q_{L_H} = q_{H_H} = \frac{p_{H_H} - h}{\propto^* (1-\propto^*)}$, which results in $\pi_{H_H} = \pi_{L_H} = \frac{(p_{H_H} - h)^2}{\propto^* (1-\propto^*)}$.

Finally, a firm offering certified high quality selects the quantity supplied that maximizes individual profits as follows:

$\max_{q_{H_C}} = q_{H_C}(p_{H_C} - h - c)$, where $p_{H_C} = 1 - \propto^* + p_{H_H} - (1 - \propto^*) Q_C$.

Since all firms selecting this strategy are equal, the first order condition of the maximization problem can be written as $q_{H_C} = \frac{1-\propto^* + p_{H_H} - h - c}{(1-\propto^*)(n_{H_C}+1)}$. Substituting this into the corresponding aggregate demand function leads to the best response price $p_{H_C} = \frac{1-\propto^* + p_{H_H} + (h+c)n_{H_C}}{n_{H_C}+1}$. This results in $q_{H_C} = \frac{p_{H_C} - h - c}{1-\propto^*}$ and $\pi_{H_C} = \frac{\left(p_{H_C} - h - c\right)^2}{1-\propto^*}$.

The respective numbers of firms in equilibrium are obtained by imposing the condition $\pi_{L_L} = \pi_{L_H} = \pi_{H_H} = \pi_{H_C}$. Since $\pi_{L_L} = \frac{p_{L_L}^2}{\propto^*}$, $\pi_{H_H} = \pi_{L_H} = \frac{\left(p_{H_H} - h\right)^2}{\propto^*(1-\propto^*)}$ and $\pi_{H_C} = \frac{\left(p_{H_C} - h - c\right)^2}{1-\propto^*}$, the conditions $p_{L_H} = p_{H_H} = (1-\propto^*)^{1/2} p_{L_L} + h$, $p_{H_C} = \left(\frac{1-\propto^*}{\propto^*}\right)^{1/2} p_{L_L} + h + c$ and $n_{L_L} + n_{L_H} + n_{H_H} + n_{H_C} = N$ must hold in equilibrium.

We use the best responses $p_{L_L} = \frac{p_{H_H}}{n_{L_L}+1}$, $p_{H_H} = p_{L_H} = \frac{\propto^* p_{H_C} + (1-\propto^*) p_{L_L} + h(n_{L_H} + n_{H_H})}{n_{L_H} + n_{H_H} + 1}$ and $p_{H_C} = \frac{1-\propto^* + p_{H_H} + (h+c)n_{H_C}}{n_{H_C}+1}$, combined with the three equilibrium conditions $p_{L_H} = p_{H_H} = (1-\propto^*)^{1/2} p_{L_L} + h$, $p_{H_C} = \left(\frac{1-\propto^*}{\propto^*}\right)^{1/2} p_{L_L} + h + c$, $n_{L_L} + n_{L_H} + n_{H_H} + n_{H_C} = N$ and $n_{H_H} = \propto^* (n_{L_H} + n_{H_H})$ to obtain the corresponding equilibrium prices an numbers of firms following each strategy.

First, combining $p_{L_L} = \frac{p_{H_H}}{n_{L_L}+1}$ and $p_{L_H} = p_{H_H} = (1-\propto^*)^{1/2} p_{L_L} + h$, results in $p_{L_L} = \frac{h}{n_{L_L} + 1 - (1-\propto^*)^{1/2}}$, which leads to $n_{L_L} = \frac{h}{p_{L_L}} + (1-\propto^*)^{1/2} - 1$.

Second, combining $p_{H_H} = p_{L_H} = \frac{\propto^* p_{H_C} + (1-\propto^*) p_{L_L} + h(n_{L_H} + n_{H_H})}{n_{L_H} + n_{H_H} + 1}$, $p_{L_H} = p_{H_H} = (1 - \propto^*)^{1/2} p_{L_L} + h$ and $p_{H_C} = \left(\frac{1-\propto^*}{\propto^*}\right)^{1/2} p_{L_L} + h + c$ results in $p_{L_L} = \frac{\propto^*(h+c)-h}{(1-\propto^*)^{1/2}(n_{L_H}+n_{H_H}+1)-(\propto^*)^{1/2}(1-\propto^*)^{1/2}-(1-\propto^*)}$, which leads to $n_{L_H} + n_{H_H} = \frac{\propto^*(h+c)-h}{p_{L_L}(1-\propto^*)^{1/2}} + (\propto^*)^{1/2} + (1 - \propto^*)^{1/2} - 1$.

Finally, combining $p_{H_C} = \frac{1-\propto^* + p_{H_H} + (h+c) n_{H_C}}{n_{H_C}+1}$ with $p_{L_H} = p_{H_H} = (1 - \propto^*)^{1/2} p_{L_L} + h$ and $p_{H_C} = \left(\frac{1-\propto^*}{\propto^*}\right)^{1/2} p_{L_L} + h + c$, results in$p_{L_L} = \left(\frac{\propto^*}{1-\propto^*}\right)^{1/2} \frac{1-\propto^*-c}{n_{H_C}+1-(\propto^*)^{1/2}}$, which leads to $n_{H_C} = \frac{(\propto^*)^{1/2}(1-\propto^*-c)}{p_{L_L}(1-\propto^*)^{1/2}} + (\propto^*)^{1/2} - 1$.

Now, by imposing the condition $n_{L_L} + n_{L_H} + n_{H_H} + n_{H_C} = N$, the *equilibrium* price of the low quality version is $p_{L_L} = \frac{h\left[(1-\propto^*)^{1/2} + \propto^* - 1\right] + c\left[\propto^* - (\propto^*)^{1/2}\right] + (\propto^*)^{1/2}(1-\propto^*)}{(1-\propto^*)^{1/2} \cdot \left\{N + 3 - 2\left[(1-\propto^*)^{1/2} + (\propto^*)^{1/2}\right]\right\}}$, and the remaining equilibrium prices and number of firms following each strategy can be easily obtained substituting this expression in the above equations.

(iv). Only uncertified high quality is offered in equilibrium. In this case, given consumers' equilibrium belief $\propto^* \in (0,1)$, the same proportion $\propto^*$of firms labeling the product as of uncertified high quality are truly producing high quality. The equilibrium price in this case is the same as the one obtained in the following case, since this case is a limiting situation of coexistence of uncertified and certified high quality, when the mumber of firms following the latter strategy tends to zero.

(v). Only uncertified and high quality are offered in equilibrium. In this case, given consumers' equilibrium belief $\propto^* \in (0,1)$, the same proportion $\propto^*$of firms labeling the product as of uncertified high quality are truly producing high quality. The

corresponding expressions for the expected consumer's surpluses for uncertified and certified high quality are $\underline{\theta} + \alpha^* \theta_j - p_{H_H}$;; $\underline{\theta} + \theta_j - p_{H_C}$, from which the aggregate demand for uncertified high quality is $Q_H = \frac{p_{H_C} - p_{H_H}}{1-\alpha^*}$, which results in the best response price for uncertified high quality $p_{L_H} = p_{H_H} = \frac{p_{H_C} + h(n_{L_H} + n_{H_H})}{n_{L_H} + n_{H_H} + 1}$. Then, the aggregate demand for certified high quality is $Q_{H_C} = 1 - \frac{p_{H_C} - p_{H_H}}{1-\alpha^*}$, which results in the best response price $p_{H_C} = \frac{1 - \propto^* + p_{H_H} + n_{H_C}(h+c)}{n_{H_C}+1}$. Combining both reaction functions, we obtain $p_{L_H} = p_{H_H} = \frac{1 - \propto^* + n_{H_C}(h+c) + h(n_{L_H} + n_{H_H})(n_{H_C}+1)}{(n_{L_H} + n_{H_H} + 1)(n_{H_C}+1) - 1}$; $p_{H_C} = \frac{(1-\propto^*)(n_{L_H} + n_{H_H} + 1) + n_{H_C}(n_{L_H} + n_{H_H} + 1)(h+c) + h(n_{L_H} + n_{H_H})}{(n_{L_H} + n_{H_H} + 1)(n_{H_C}+1) - 1}$, which results in the corresponding profit functions $\pi_{L_H} = \pi_{H_H} = \frac{\left[1 - \propto^* + c n_{H_C}\right]^2}{(1-\propto^*)\left[(n_{L_H} + n_{H_H} + 1)(n_{H_C}+1) - 1\right]^2}$; $\pi_{H_C} = \frac{\left[(1-\propto^*)(n_{L_H} + n_{H_H} + 1) - c(n_{L_H} + n_{H_H})\right]^2}{(1-\propto^*)\left[(n_{L_H} + n_{H_H} + 1)(n_{H_C}+1) - 1\right]^2}$.

Imposing the equilibrium condition $\pi_{L_H} = \pi_{H_H} = \pi_{H_C}$ and using the fact that $n_{L_H} + n_{H_H} + n_{H_C} = N$ and $n_{H_H} = \propto^* (n_{L_H} + n_{H_H})$ we obtain that the equilibrium numbers of firms following each strategy are $n_{L_H} = cN$, $n_{H_H} = \frac{\propto^* cN}{1-\propto^*}$ *and* $n_{H_C} = \frac{(1-\propto^*-c)N}{1-\propto^*}$, which result in the equilibrium prices $p_{L_H} = p_{H_H} = \frac{1-\propto^*}{N} + h, p_{H_C} = \frac{1-\propto^*}{N} + h + c$.

Existence of the different equilibrium configurations of Proposition 3.

Here, we explore the conditions for the existence of each type of equilibrium characterized in Proposition 3 (and illustrated in Figure 1).

We first explore the conditions for the existence of the $L_L L_H H_H H_C$ equilibrium (case iii in Proposition 3). These conditions derive from the relative positions of the

indifferent consumers between purchasing low and uncertified high quality on the one hand, and uncertified and certified high quality on the other hand, in the interval [0,1]. That is, the existence of this equilibrium is guaranteed as long as there exists a subset of consumers purchasing each variety of the product. Thus, this equilibrium exists whenever $0 < \frac{p_{H_H} - p_{L_L}}{\propto^*} < \frac{p_{H_C} - p_{H_H}}{1-\propto^*} < 1$.

We first analyze the condition $0 < \frac{p_{H_H} - p_{L_L}}{\propto^*}$. Substituting the equilibrium values for p_{L_L} and p_{H_H} presented in part (iii) of Proposition 3, we obtain:

$$h\left\{1 - \frac{\left[1-(1-\propto^*)^{1/2}\right]^2}{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]}\right\} + c\left\{\frac{\left[(1-\propto^*)^{1/2}-1\right][\propto^*-(\propto^*)^{1/2}]}{(1-\propto^*)^{1/2}\cdot\left\{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]\right\}}\right\} +$$

$$\frac{\left[(1-\propto^*)^{1/2}-1\right](\propto^*)^{1/2}(1-\propto^*)^{1/2}}{\left\{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]\right\}} > 0.$$

The term that accompanies h is positive and the term that accompanies c is also positive. Under equality, this condition then defines a linear negative relationship between h and c, namely $\theta^{LH} = 0$ in Figure 1. The vertical intercept of this linear relationship (i.e., the value of c when $h = 0$) is $\frac{1-\propto^*}{1-(\propto^*)^{1/2}} > 0$. Note that this condition is not satisfied for h and c sufficiently small, in particular, it is not satisfied for $h = c = 0$.

Second, we analyze the condition $\frac{p_{H_C} - p_{H_H}}{1-\propto^*} < 1$. Substituting the corresponding equilibrium expressions for p_{L_L}, $p_{H_H}$and p_{H_C} presented in part (iii) of Proposition 3, we obtain the condition:

$$h\left\{\frac{\left[\frac{1}{(\propto^*)^{1/2}} - 1\right]\left[(1-\propto^*)^{1/2} + \propto^* - 1\right]}{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]}\right\} + c\left\{\frac{\left[\frac{1}{(\propto^*)^{1/2}} - 1\right]\left[\propto^* - (\propto^*)^{1/2}\right]}{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]} + 1\right\}$$

$$< (1-\propto^*)\left\{1 - \frac{\left[1-(\propto^*)^{1/2}\right]}{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]}\right\}$$

Under equality, a negative linear relationship between h and c is established, since the term that accompanies h is clearly positive, while the term that accompanies c is positive as well (the first term of the expression in brackets is negative but smaller than one in absolute value). This linear relationship is represented by the label $\theta^{HC} = 1$ in Figure 1. The vertical intercept of this relationship (i.e., the value of c when $h = 0$) in this case is $\frac{\left\{1-\frac{1-(\alpha^*)^{1/2}}{N+3-2\left[(1-\alpha^*)^{1/2}+(\alpha^*)^{1/2}\right]}\right\}(1-\alpha^*)}{\left\{1-\frac{\left(1-(\alpha^*)^{1/2}\right)^2}{N+3-2\left[(1-\alpha^*)^{1/2}+(\alpha^*)^{1/2}\right]}\right\}} > \frac{1-\alpha^*}{1-(\alpha^*)^{1/2}}$.

Finally, condition $\frac{p_{H_H}-p_{L_L}}{\alpha^*} < \frac{p_{H_C}-p_{H_H}}{1-\alpha^*}$ results in:

$$\frac{h}{\alpha^*}\left\{\frac{\left[1-(1-\alpha^*)^{1/2}-(\alpha^*)^{1/2}\right]\left[(1-\alpha^*)^{1/2}+\alpha^*-1\right]}{(1-\alpha^*)\{N+3-2[(1-\alpha^*)^{1/2}+(\alpha^*)^{1/2}]\}}+1\right\}$$

$$+\frac{c}{1-\alpha^*}\left\{\frac{\left[1-(1-\alpha^*)^{1/2}-(\alpha^*)^{1/2}\right]\left[\alpha^*-(\alpha^*)^{1/2}\right]}{\alpha^*\{N+3-2[(1-\alpha^*)^{1/2}+(\alpha^*)^{1/2}]\}}-1\right\}$$

$$+\frac{\left[1-(1-\alpha^*)^{1/2}-(\alpha^*)^{1/2}\right]}{(\alpha^*)^{1/2}\{N+3-2[(1-\alpha^*)^{1/2}+(\alpha^*)^{1/2}]\}} < 0$$

Under equality, a positive linear relationship is established between h and c (which is labeled as $\theta^{LH} = \theta^{HC}$ in Figure 1), since the term that accompanies h is positive (the first term in brackets is negative, but it is surely lower than one in absolute value), while the term that accompanies c is negative as well (the first term in brackets is negative, and it is surely smaller than one in absolute value).

As a result of the combination of the three conditions $\theta^{LH} > 0$, $\theta^{HC} < 1$ and $\theta^{LH} < \theta^{HC}$, for any $\alpha^* \in (0,1)$, there always exist a set of parameters for (h, c) such that equilibrium $L_L L_H H_H H_C$ exists. The size of this range of parameters depends on consumers' equilibrium belief, α^*, and the total number of firms. When one or more of

the three conditions for the existence of this equilibrium are not satisfied, the system results in any of the other four possibilities stated in Proposition 3.

For example, if $\theta^{HC} \geq 1$, equilibrium $L_L L_H H_H$ emerges, see Figure 1. The equilibrium prices and the number of firms in equilibrium in this case are presented in part (ii) of Proposition 3. Thus, for this equilibrium to exist, condition $\theta^{HC} \geq 1$ needs to hold and also condition $0 < n_i < N$, i $= L_L, L_H, H_H$, which is satisfied as long as $h < \propto^*$. Otherwise, only the low quality version of the product is offered in the market. Thus, conditions $\theta^{HC} \geq 1$ and $h < \propto^*$ jointly characterize the existence of equilibrium $L_L L_H H_H$.

When $\theta^{LH} \geq \theta^{HC}$, only the low and the certified high quality versions of the product can coexist in equilibrium (equilibrium $L_L H_C$, see Figure 1), as characterized in part (i) of Result 3. Therefore, for this equilibrium to exist, condition $\theta^{LH} \geq \theta^{HC}$ needs to hold and also condition $0 < n_i < N$, i $= L_L, H_C$, which is satisfied as long as $h + c < 1$. Otherwise, only the low quality version of the product is offered in the market. Therefore, conditions $\theta^{LH} \geq \theta^{HC}$ and $h + c < 1$ jointly characterize the existence of equilibrium $L_L H_C$.

Finally, when $\theta^{LH} \leq 0$, only the uncertified and the certified high quality versions of the product can coexist (equilibria $L_H H_H$ and $L_H H_H H_C$, see Figure 1), with corresponding equilibrium characteristics given in parts (iv) and (v) of Proposition 3, respectively. The existence of equilibrium $L_H H_H H_C$ is guaranteed as long as $\theta^{LH} \leq 0$ and $c < 1 - \propto^*$. By contrast, equilibrium $L_H H_H$ emerges if $\theta^{LH} \leq 0$ and $c < 1 - \propto^*$.

Proof of Result 3. Starting with p_{L_L}, note that $p_{L_L} = \frac{1}{N}$ under both large and low fines for fraud (see Propositions 1 and 2, respectively), and also in case (i) of Proposition 3.

In cases (ii) and (iii) of Proposition 3 (the only additional cases where the low quality version of the product is offered, we respectively have $p_{L_L} = \frac{\propto^*}{N} < \frac{1}{N}$ since $\propto^* \in (0,1)$, and $p_{L_L} = \frac{h\left[(1-\propto^*)^{1/2}+\propto^*-1\right]+c\left[\propto^*-(\propto^*)^{1/2}\right]+(\propto^*)^{1/2}(1-\propto^*)}{(1-\propto^*)^{1/2}\cdot\{N+3-2\left[(1-\propto^*)^{1/2}+(\propto^*)^{1/2}\right]\}} < \frac{1}{N}$, since $h < \propto^* < (\propto^*)^{1/2}$ in area $L_L L_H H_H H_C$ (see Figure 1). Note that in cases (iv) and (v), the low quality version is produced but marketed as high quality. Nevertheless, $p_{L_H} = \frac{1-\propto^*}{N} + h < \frac{1}{N}$ as well, since $N > 1$ and $h < \propto^*$ in areas $L_H H_H$ and $L_H H_H H_C$ (see Figure 1).

Considering now p_{H_H}, note that $p_{H_H} = \frac{1}{N} + h$ in the case of high institutional quality (Proposition 1). If we observe the cases where this price arises (cases ii-v of Proposition 3), we then have $p_{H_H} = \frac{\propto^*}{N} + h < \frac{1}{N} + h$ (case ii); $p_{H_H} = (1-\propto^*)^{1/2} p_{L_L} + h < (1-\propto^*)^{1/2}\frac{1}{N} + h < \frac{1}{N} + h$ (case iii); and $p_{H_H} = \frac{1-\propto^*}{N} + h < \frac{1}{N} + h$ (cases iv and v).

Finally, regarding p_{H_C}, note that $p_{H_C} = \frac{1}{N} + h + c$ in the case of low fines (Proposition 2). If we now look at the cases where this price arises (cases i, iii and v of Proposition 3), we then have $p_{H_C} = \frac{1}{N} + h + c$ (case i), $p_{H_C} = \left(\frac{1-\propto^*}{\propto^*}\right)^{1/2} p_{L_L} + h + c < \frac{1}{N} + h + c$ (case iii), since $h < \propto^* < (\propto^*)^{1/2}$ in area $L_L L_H H_H H_C$ (see Figure 1), and $p_{H_C} = \frac{1-\propto^*}{N} + h + c < \frac{1}{N} + h + c$ (case v).

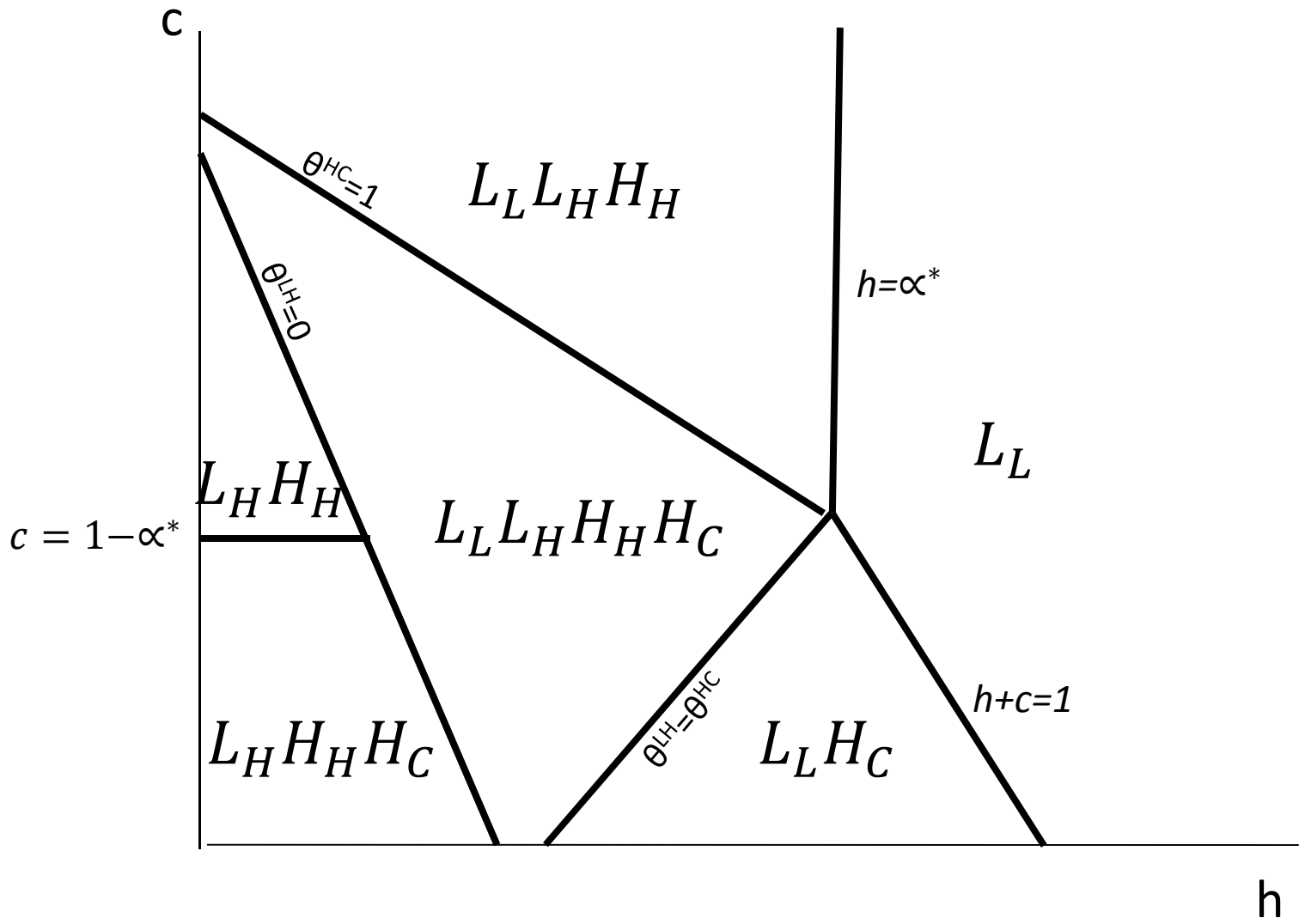


Figure 1. Equilibrium configurations for intermediate fines for fraud and small *N*

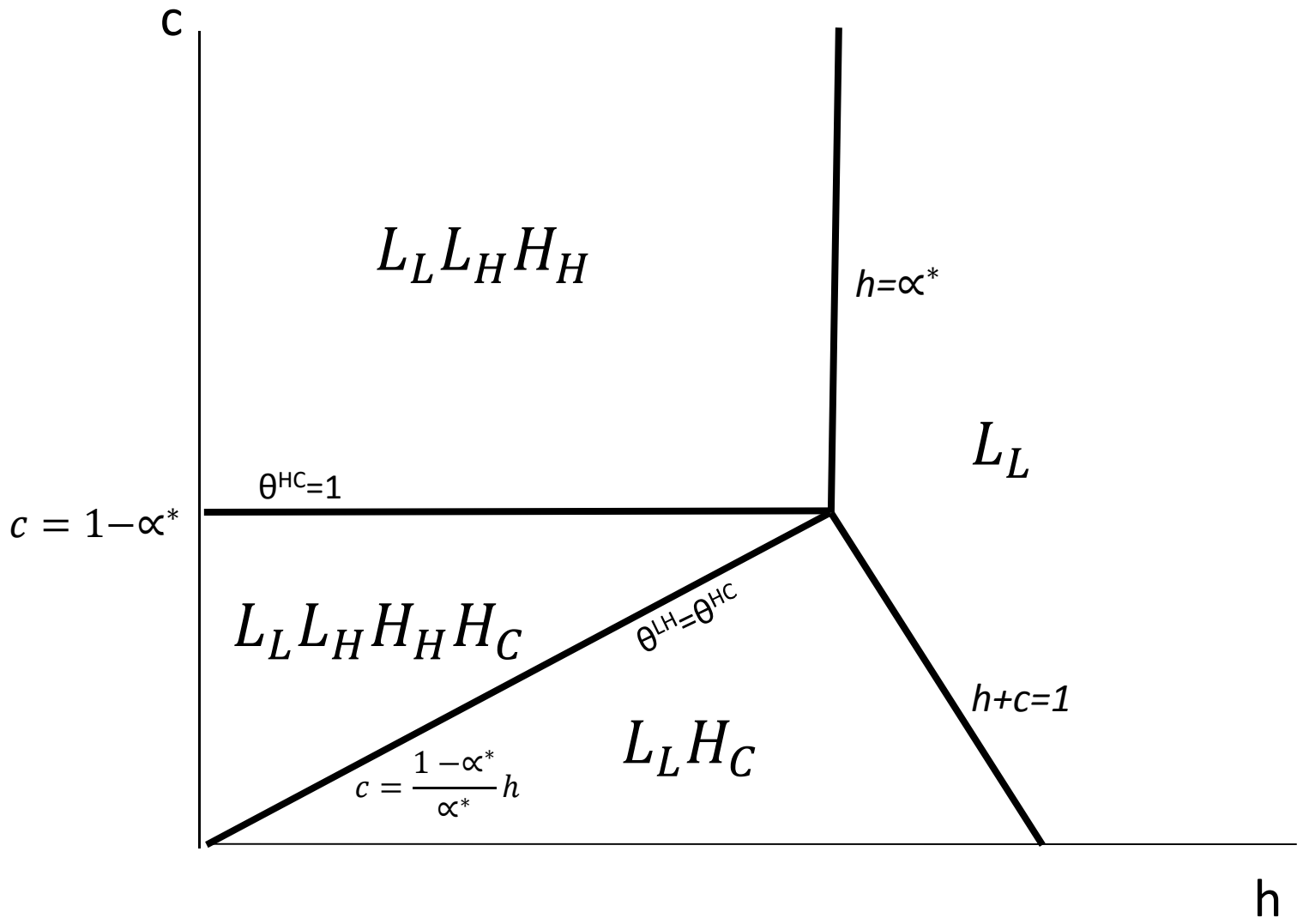
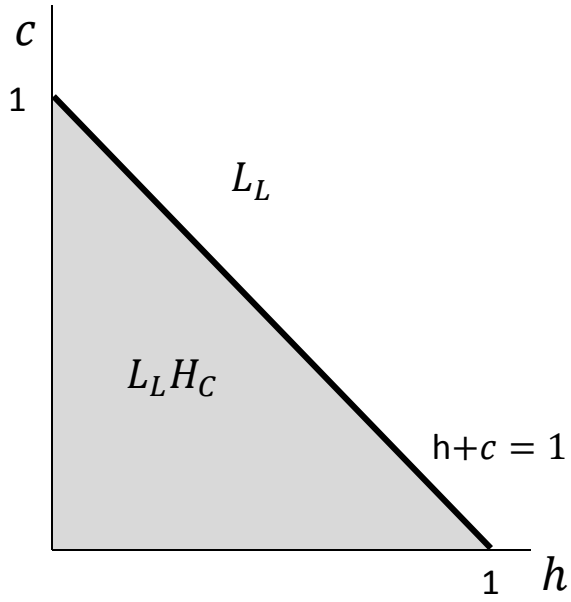
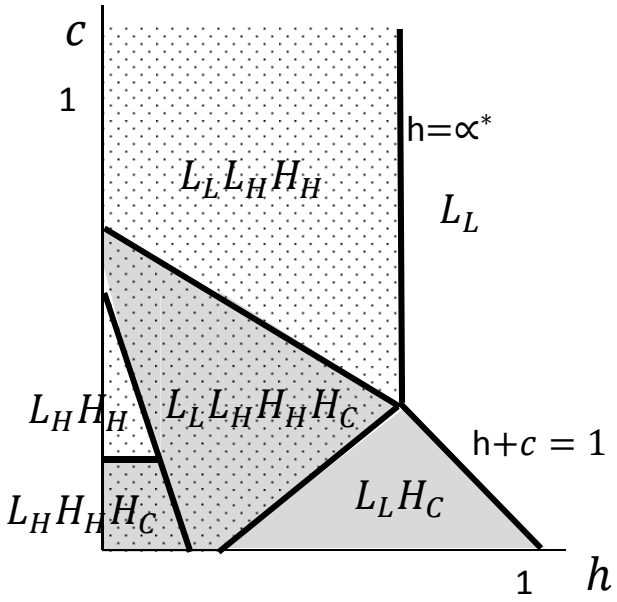


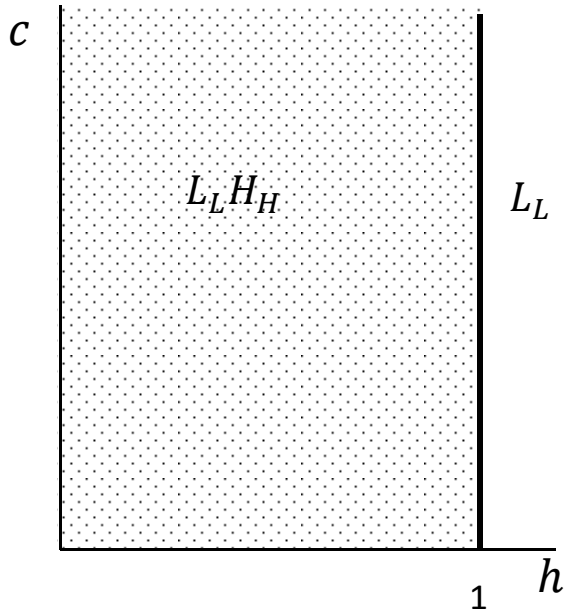
Figure 2. Equilibrium configurations for intermediate fines for fraud and large *N*



3a) Low fines

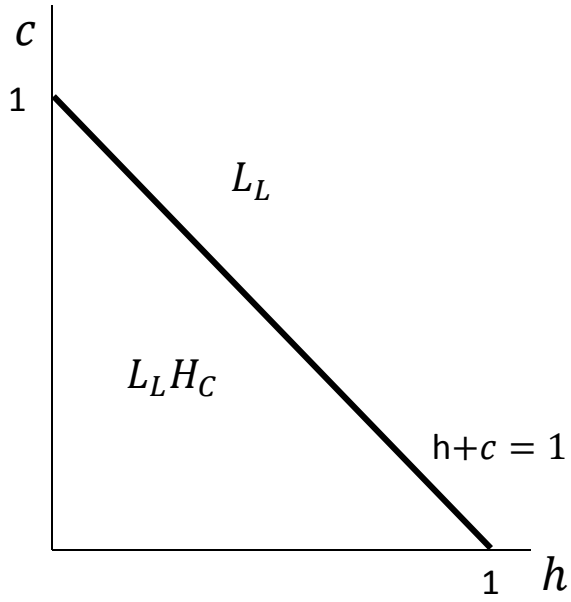


3b) Intermediate fines

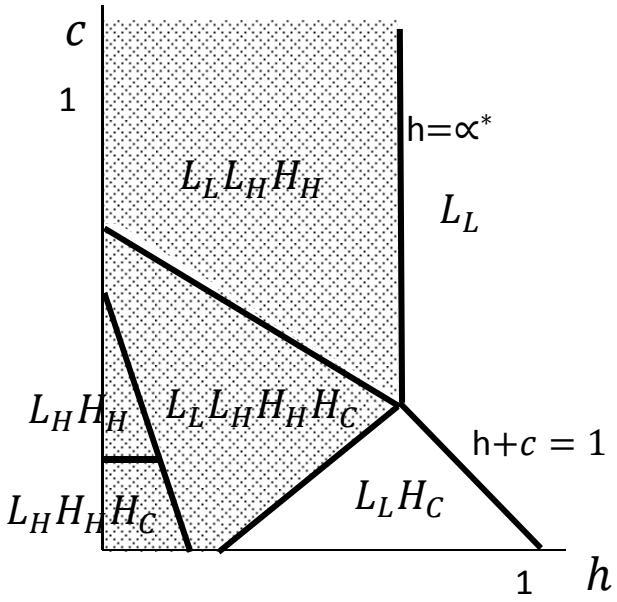


3c) Large fines

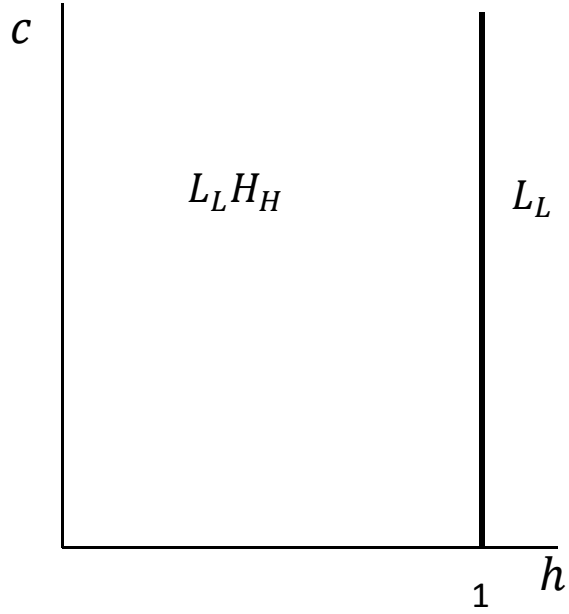
Figure 3. Equilibrium configurations for different fines for fraud and small *N*. Grey areas represent equilibria with certification. Dotted areas represent the presence of uncertified high quality.



4a) Low fines



4b) Intermediate fines



4c) Large fines

Figure 4. Equilibrium configurations for different fines for fraud and small *N*. Shadowed areas represent equilibria with fraud.