Endogenous Antitrust Enforcement in the Presence of a Corporate Leniency Program*

Joseph E. Harrington, Jr.
Department of Business Economics & Public Policy
The Wharton School
University of Pennsylvania
3620 Locust Walk
Philadelphia, PA 19102
harrij@wharton.upenn.edu
(Visiting: Departamento de Economía, Universidad Carlos III de Madrid)

Myong-Hun Chang
Department of Economics
Cleveland State University
Cleveland, OH 44115
m.chang@csuohio.edu

Incomplete and not for general circulation
24 September 2012

*This is a heavily revised version of "The Impact of a Corporate Leniency Program on Antitrust Enforcement and Cartelization" (April 2010). For their valuable comments on that paper, we thank Louis Kaplow, Kai-Uwe Kühn, Yossi Spiegel, Juuso Välimäki, participants of the 15th WZB Conference on Markets and Politics/2nd Conference of the Research Network on Innovation and Competition Policy, Helsinki Center of Economic Research Conference on Cartels and Collusion, Tercer Taller de Organización Industrial (Zapallar, Chile), Third Annual Searle Research Symposium on Antitrust Economics and Competition Policy (Northwestern University) and seminar participants at George Washington University, Lafayette College, University of Texas at Dallas, University of North Carolina at Chapel Hill, and Northwestern University (Department of Management & Strategy - Kellogg Graduate School of Management). The first author gratefully acknowledges the support of the National Science Foundation (SES-0516943).

Abstract: Constructing a birth and death model of cartels, this paper examines the impact of instituting a corporate leniency program on the steady-state frequency of cartels in a population of industries. An innovative feature of the model is taking account of how a leniency program impacts the effectiveness of a competition authority in prosecuting cases without a leniency applicant. It is shown that a leniency program is assured of lowering the cartel rate when leniency cases take up sufficiently few competition authority resources or when enforcement was initially very weak. When leniency cases are just as intensive to prosecute and penalties are sufficiently low then a leniency program is not only ineffective but actually *raises* the cartel rate because of its deleterious effect on non-leniency enforcement. Measuring the performance of a leniency program using the number of leniency applications is shown to be problematic because a leniency program can lower the cartel rate while generating no applications and raise the cartel rate while generating many applications.

1 Introduction

The 1993 revision of the Corporate Leniency Program of the U.S. Department of Justice's Antitrust Division gives a member of a cartel the opportunity to avoid government penalties if it is the first to fully cooperate and provide evidence. In the U.S., this program is arguably the most significant policy development in the fight against cartels since the Clayton Act instituted private treble damages in 1914. As reported by Antitrust Division officials, the leniency program is the primary generator of cartel cases, and the information provided by those admitted to the program has been instrumental in securing the convictions of other cartel members. Deputy Assistant Attorney General Scott Hammond stated in 2005:[1]

> The Antitrust Division's Corporate Leniency Program has been the Division's most effective investigative tool. Cooperation from leniency applicants has cracked more cartels than all other tools at our disposal combined.

The widespread usage of the leniency program in the U.S. soon led to the adoption of similar programs in other countries. In 1996, the European Commission (EC) instituted a leniency program and a decade later 24 out of 27 EU members had one. Today, leniency programs span the globe from Canada to the United Kingdom to Japan to South Africa to Brazil. Since 1995, more than 20% of discovered international cartels have been awarded amnesty by at least one competition authority (Connor, 2008). The EC provided partial or full leniency in 45 of 50 cartel cases decided during 1998-2007, and leniency lowered average fines per cartel by almost 40% (Veljanovski, 2007). In sum, leniency programs have been widely introduced and utilized throughout the world and are now present in more than 50 countries and jurisdictions.[2]

In light of the widespread adoption and usage of leniency programs, a considerable body of scholarly work has developed to understand these programs and assess how they can be better designed; a review of some of this research is provided in Spagnolo (2008). Starting with the pioneering paper of Motta and Polo (2003), there has been a sequence of theoretical analyses including Spagnolo (2003), Aubert, Kovacic, and Rey (2006), Chen and Rey (2007), Harrington (2008), and Choi and Gerlach (2010). While models and results vary, the overall conclusion is that leniency programs make collusion more difficult.[3] There is also a growing body of experimental work which similarly provides evidence of the efficacy of leniency programs including Apesteguia, Dufwenberg, and Selten (2007), Hinloopen and Soetevent (2008), Dijkstra, Haan, and Schoonbeek (2011), and Bigoni et al (2012). These experimental studies generally find

[1] Scott D. Hammond, "Cracking Cartels With Leniency Programs," OECD Competition Committee, Paris, France, October 18, 2005.

[2] Borrell, Jiménez, and García (2012) estimates how leniency programs have changed the perceptions of managers throughout the world.

[3] There are a variety of effects at work when a leniency program is put in place and some serve to make collusion easier; see Ellis and Wilson (2001) and Chen and Harrington (2007). Generally, these effects net out so that fewer cartels form when there is a leniency program.

that a leniency program reduces cartel formation though some studies also find that prices are higher, conditional on a cartel forming, when there is a leniency program. Finally, there are an increasing number of empirical studies that measure the impact of leniency programs. Using data over 1985-2005 for the United States, Miller (2009) finds evidence that the 1993 revision reduced the latent cartel rate. In contrast, Brenner (2009) does not find evidence that collusion was made more difficult with the European Commission's 1996 Corporate Leniency Program, though his data is for 1990-2003 and thus does not encompass an important revision in the program in 2002. Preliminary findings in Klein (2010) and Zhou (2011) suggest that the EC's leniency program has been effective.

While the empirical evidence is mixed, the general conclusion from theoretical and experimental research is that leniency programs are effective in shutting down cartels and deterring cartel formation. However, those findings were derived under a crucial but problematic assumption that non-leniency enforcement is unaffected by the introduction of a leniency program. More specifically, it is assumed that the probability that a cartel is discovered, prosecuted, and convicted - in the absence of a firm coming forward under the leniency program - is unchanged with the adoption of a leniency program. Not only is that assumption almost certain to be violated, but conclusions about the efficacy of a leniency program could significantly change once this probability is made endogenous. Let us argue both points.

With the introduction of a leniency program, the investigation of cases not involving leniency is likely to change and, as a result, this will affect the probability that a cartel is caught and convicted. As a competition authority has limited resources, if resources are used to handle leniency cases then fewer resources are available to effectively prosecute non-leniency cases. This doesn't necessarily imply that non-leniency enforcement is weaker, however. If a leniency program is successful in reducing the number of cartels, there will be fewer non-leniency cartel cases, in which case the authority may still have ample resources to effectively prosecute them. Furthermore, an optimizing competition authority is likely to adjust its enforcement policy - for example, how it allocates prosecutorial resources across cases - in response to what is occurring with leniency applications. Thus, while we expect the probability that a cartel is caught and convicted to change when a leniency program is put in place, it isn't clear in which direction it will go.

The next point to note is that a change in the likelihood of getting a conviction for a non-leniency case has implications for the efficacy of the leniency process itself. A cartel member will apply for leniency only if it believes that doing so is better than running the risk of being caught and paying full penalties. Thus, the probability of being caught and convicted is integral to inducing firms to apply for leniency. If this probability is very low then no cartel member will use the leniency program, while if the probability is sufficiently high then, under the right circumstances, a cartel member will apply for leniency. The efficacy of a leniency program is then intrinsically tied to how a leniency program affects the probability of being caught and convicted when no firm applies for leniency.

These issues have been recognized in the policy realm as legitimate concerns. In recent years the Directorate General Competition (DG Comp) of the European Commission has been overwhelmed with leniency applications which limits the amount of resources for prosecuting other cases:[4]

> DG Competition is now in many ways the victim of its own success; leniency applicants are flowing through the door of its Rue Joseph II offices, and as a result the small Cartel Directorate is overwhelmed with work. ... It is open to question whether a Cartel Directorate consisting of only approximately 60 staff is really sufficient for the Commission to tackle the 50 cartels now on its books.

Furthermore, the interaction between adoption of a leniency program and enforcement through means other than leniency applications is emphasized in Friederiszick and Maier-Rigaud (2008). Both authors were members of DG Comp and their paper argues that the DG Comp should be more active in detecting cartels and more generally in initiating cases *because* of the success of the leniency program.

The primary contribution of this paper is to assess the impact of a leniency program on the cartel rate while endogenizing the efficacy of non-leniency enforcement. This analysis is done in the context of a population of industries in which cartels are formed and dissolved either due to internal collapse or the efforts of the competition authority. The focus is on how a leniency program affects the steady-state fraction of industries that are cartelized. Non-leniency enforcement is measured by the probability that a cartel is caught, prosecuted, and convicted without use of the leniency program. The probability of conviction depends on the size of the competition authority's caseload which is composed of both leniency and non-leniency cases. This structure creates a feedback relationship that simultaneously determines the efficacy of a leniency program and non-leniency enforcement: A leniency program affects the number of cartels which influences the competition authority's caseload which influences the rate at which non-leniency cases are won which influences expected penalties which influences cartel formation and expected cartel duration and, therefore, the number of cartels. It will be shown that allowing non-leniency enforcement to respond can either reinforce the efficacy of a leniency program or work against it, even to the point that the cartel rate is higher after the introduction of a leniency program.

In the next section, the model is presented. In Section 3, the conditions determining the equilibrium cartel rate are derived. Existence and some basic properties of the equilibrium cartel rate are established in Section 4. Sections 5 and 6 provide the central results. In Section 5, sufficient conditions are derived for a leniency program to lower the cartel rate and for a leniency program to raise the cartel rate. The section concludes with a general discussion of what is required for a leniency program to be effective.

[4] Riley (2007), pp. 1-2.

2 Model

The modelling strategy is to build a birth and death Markov process for cartels in order to generate an average cartel rate for a population of industries, and to then assess how the introduction of a leniency program influences the frequency of cartels. We build upon the birth and death process developed in Harrington and Chang (2009) by allowing for a leniency program and, most crucially, endogenizing the probability that a cartel is convicted through non-leniency enforcement.[5]

2.1 Industry Environment

Firm behavior is modelled using a modification of a Prisoners' Dilemma formulation. Firms simultaneously decide whether to *collude* (set a high price) or *compete* (set a low price). Prior to making that choice, firms observe a stochastic realization of the market's profitability that is summarized by the variable $\pi \geq 0$.[6] If all firms choose *collude* then each firm earns π, while if all choose *compete* then each earns $\alpha\pi$ where $\alpha \in [0,1)$. $1-\alpha$ then measures the competitiveness of the non-collusive environment. π has a continuously differentiable cdf $H : [\underline{\pi}, \overline{\pi}] \rightarrow [0,1]$ where $0 < \underline{\pi} < \overline{\pi}$. $h(\cdot)$ denotes the associated density function and let $\mu \equiv \int \pi h(\pi) d\pi$ denote its finite mean. If all other firms choose *collude*, the profit a firm earns by deviating - choosing *compete* - is $\eta\pi$ where $\eta > 1$. This information is summarized in the table below. Note that the Bertrand price game is represented by $(\alpha, \eta) = (0, n)$ where n is the number of firms. The Cournot quantity game with linear demand and cost functions in which firms collude at the joint profit maximum is represented by $(\alpha, \eta) = \left(\frac{4n}{(n+1)^2}, \frac{(n+1)^2}{4n}\right)$.[7]

Own action	All other firms' actions	Own profit
collude	*collude*	π
compete	*collude*	$\eta\pi$
compete	*compete*	$\alpha\pi$

Firms interact in an infinite horizon setting where $\delta \in (0,1)$ is the common discount factor. It is not a repeated game because, as explained later, each industry is in one of two states: cartel and non-cartel. If *firms are a cartel* then they effectively collude only when it is incentive compatible. More specifically, if firms are cartelized then they simultaneously choose between *collude* and *compete*, and, at the same time, whether or not to apply to the corporate leniency program. Details on the description of the leniency program are provided later. If it is incentive compatible for all firms to choose *collude* then each earns π. If instead a firm prefers *compete* when all other firms choose *collude* then collusion is not incentive compatible (that is, it is not part

[5] The Harrington-Chang birth and death process for cartels is estimated in Hyytinen, Steen, and Toivanen (2010) who use data from the Finnish cartel registry.

[6] The informational setting is as in Rotemberg and Saloner (1986).

[7] We have only specified a firm's profit when all firms choose *compete*, all firms choose *collude*, and it chooses *compete* and all others firm choose *collude*. We must also assume that *compete* strictly dominates *collude* for the stage game. It is unnecessary to provide any further specification.

of the subgame perfect equilibrium for the infinite horizon game) and each firm earns $\alpha\pi$. In that case, collusion is not achieved. If *firms are not a cartel* then each firm earns $\alpha\pi$ as, according to equilibrium, they all choose *compete*.

At the end of the period, there is the random event whereby the competition authority (CA) may pursue an investigation; this can only occur if firms colluded in the current or previous period and no firm applied for leniency.[8] Let $\sigma \in [0,1]$ denote the probability that firms are discovered, prosecuted, and convicted (below, we will endogenize σ though, from the perspective of an individual industry, it is exogenous). In that event, each firm incurs a penalty of F.

It is desirable to allow F to depend on the extent of collusion. Given there is only one level of collusion in the model, the "extent of collusion" necessarily refers to the number of periods that firms had colluded. A proper accounting of that effect would require that each cartel have a state variable which is the length of collusion which would seriously complicate the analysis. As a simplifying approximation, it is instead assumed that the penalty is proportional to the average increase in profit from being cartelized (rather than the *realized* increase in profit). If Y denotes the expected per period profit from being in the "cartel state" then $F = \gamma(Y - \alpha\mu)$ where $\gamma > 0$ and $\alpha\mu$ is average non-collusive profit. This specification avoids the need for state variables but still allows the penalty to be sensitive to the (average) extent of collusion.[9]

In addition to being discovered by the CA, a cartel can be uncovered because one of its members comes forth under the corporate leniency policy. Suppose a cartel is in place. If a single firm applies for leniency then all firms are convicted for sure and the firm that applied receives a penalty of θF where $\theta \in [0,1)$, while the other cartel members each pay F. If all firms simultaneously apply for leniency then each firm pays a penalty of ωF where $\omega \in (\theta, 1)$. For example, if only one firm can receive leniency and each firm has an equal probability of being first in the door then $\omega = \frac{n-1+\theta}{n}$ when there are n cartel members. It is sufficient for the ensuing analysis that we specify the leniency program when either one firm applies or all firms apply. Also, leniency is not awarded to firms that apply after another firm has done so.

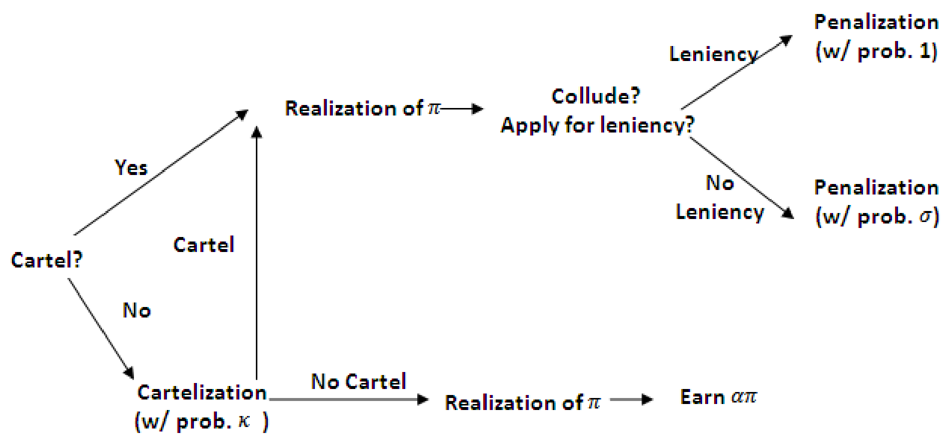
From the perspective of firms, competition policy is summarized by the four-tuple $(\sigma, \gamma, \theta, \omega)$ which are, respectively, the probability of paying penalties through non-leniency enforcement, the penalty multiple, the leniency parameter when only one firm applies (where $1-\theta$ is the proportion of fines waived), and the leniency parameter when all firms apply (where $1-\omega$ is the proportion of fines waived). As there are assumed to be only corporate penalties, our model is better suited for non-U.S. jurisdictions, such as the European Union, which lack individual penalties.

Next, let us describe how an industry's cartel status evolves. Suppose it enters the period cartelized. The industry will exit the period still being a cartel if: 1)

[8] To allow it to depend on collusion farther back in time would require introducing another state variable that would unnecessarily complicate the analysis. Having it depend on collusion in the previous period will simplify some of the expressions and, furthermore, it seems quite reasonable that detection can occur, to a limited extent, after the fact.

[9] A more standard assumption in the literature is to assume F is fixed which is certainly simpler but less realistic than our specification.

all firms chose *collude* (which requires that collusion be incentive compatible); 2) no firm applied for leniency; and 3) the CA did not discover and convict the firms of collusion. Otherwise, the cartel collapses and firms revert to the "no cartel" state. If instead the industry entered the period in the "no cartel" state then with probability $\kappa \in (0,1)$ firms cartelize. For that cartel to still be around at the end of the period, conditions (1)-(3) above must be satisfied. Note that whenever a cartel is shutdown - whether due to internal collapse, applying to the leniency program, or having been successfully prosecuted - the industry may re-cartelize in the future. Specifically, it has an opportunity to do so with probability κ in each period that it is not currently colluding.[10] The timing of events is summarized in the figure below.



In modelling a population of industries, it is compelling to allow industries to vary in terms of cartel stability. For this purpose, industries are assumed to differ in the parameter η. If one takes this assumption literally, it can be motivated by heterogeneity in the elasticity of firm demand or the number of firms (as with the Bertrand price game). Our intent is not to be literal but rather to think of this as a parsimonious way in which to encompass industry heterogeneity. Let the cdf on industry types be represented by the continuously differentiable strictly increasing function $G : [\underline{\eta}, \overline{\eta}] \rightarrow [0,1]$ where $1 < \underline{\eta} < \overline{\eta}$. $g(\cdot)$ denotes the associated density function. The appeal of η is that it is a parameter which influences the frequency of collusion but does not directly affect the value of the firm's profit stream since, in equilibrium, firms do not cheat; this property makes for an easier analysis.

2.2 Enforcement Technology

Non-leniency enforcement is represented by σ which is the probability that a cartel pays penalties without one of its members having entered the leniency program. Here,

[10] Alternatively, one could imagine having two distinct probabilities - one to reconstitute collusion after a firm cheated (the probability of moving from the punishment to the cooperative phase) and another to reform the cartel after having been convicted. For purposes of parsimony, those two probabilities are assumed to be the same.

we explain how σ is determined. σ is the compound probability that: 1) the cartel is discovered by the CA; 2) the CA decides to investigate the cartel; and 3) the CA is successful in its investigation and penalties are levied. The initial discovery of a cartel is presumed to be exogenous and to come from customers, uninvolved employees, the accidental discovery of evidence through a proposed merger, and so forth. $q \in [0, 1]$ denotes the probability of discovery and is a parameter throughout the paper. What the CA controls is how many cases to take on which is represented by $r \in [0, 1]$ which is the fraction of reported cases that the CA chooses to investigate. Initially, we will derive results when r is fixed and then allow r to be endogenous. Finally, of those cases discovered and investigated, the CA is successful in a fraction $s \in [0, 1]$ of them where s is determined by the relationship between the CA's resources and its caseload.[11]

The CA is faced with a resource constraint: the more cases it takes on, the fewer resources are applied to each case and the lower is the probability of winning any individual case. More specifically, it is assumed

$$s = p(\lambda L + R) \text{ where } \lambda \in (0, 1].$$

L is the number (or mass) of leniency cases, R is the number of non-leniency cases, and s is the proportion of R cases that result in a conviction. Leniency cases are assumed to be won for sure. $\lambda \leq 1$ because leniency cases may take up fewer resources than those cases lacking an informant. We will refer to $L + R$ as the number of cases and $\lambda L + R$ as the caseload. $p : [0, 1] \to [0, 1]$ is a continuous decreasing function so that a bigger caseload means a lower probability of winning a non-leniency case. In sum, the probability that a cartel pays penalties is

$$\sigma = q \times r \times s = q \times r \times p(\lambda L + R).$$

σ is endogenous because s is determined by the number of leniency and non-leniency cases which depends on the number of cartels, and r may be chosen by the CA.[12]

Key to the analysis is the implicit assumption that the CA faces a resource constraint in the sense that resources per case decline with the number of cases as reflected in the specification that the probability of any investigation being successful is decreasing in the caseload. In practice, an CA can move around resources to handle additional cartel activity by, for example, shifting lawyers and economists from merger cases to cartel cases. However, there is a rising opportunity cost in doing so and that ought to imply that resources per cartel case will decline with the number

[11] In a richer model, we could allow for heterogeneity across cartel cases in terms of the perceived difficulty of gaining a conviction; that is, s is cartel-specific. The CA would then decide not only how many cases to pursue but which cases to pursue.

[12] It should be noted that Motta and Polo (2003) do allow for optimal enforcement expenditure by modelling a trade-off between monitoring and prosecution. They endow a CA with a fixed amount of resources that can be allocated between finding suspected episodes of collusion and prosecuting the cases that are found or, in the language of our model, between raising q and lowering s (assuming $r = 1$). However, the model is very different from ours - for example, they do not consider a population of industries and do not solve for the steady-state frequency of cartels - and it does not address the questions we are raising here.

of cartel cases. Of course, CA officials can lobby their superiors (either higher level bureaucrats or elected officials) for a bigger budget but, at least in the U.S., the reality is that the CA's budget does not scale up with its caseload. While the budget of the Antitrust Division of the U.S. Department of Justice is increasing in GDP (Kwoka, 1999), DOJ antitrust case activity is actually countercyclical (Ghosal and Gallo, 2001) which substantiates our assumption that resources per case is declining in the number of cases that a CA takes on.

The last element to specify is the determination of the fraction of cases that the CA takes on, r, which requires specifying an objective to the CA. As a benchmark, it is assumed here that the CA is welfare-maximizing in that it chooses r to minimize the cartel rate. The results in Section 5 actually are true either holding r fixed or with this objective assigned to the CA. The numerical results in Section 6 assume r is selected to minimize the cartel rate because it eliminates having to specify a value for one more parameter. We view the assumption that the CA minimizes the cartel rate to be a useful benchmark for gaining insight into the possible implications of a leniency program though not necessarily as a good description of CA behavior. However a CA is rewarded, it is natural to assume that rewards are based on observable measures of performance. Given that the cartel rate is not observable (only discovered cartels are observable) then presumably the CA is not rewarded based on how its policies affect the cartel rate (at least not directly). Thus, it is not clear that there is an incentive scheme that will induce the CA to minimize the cartel rate. We intend to explore the modelling of the CA in future research.[13]

3 Equilibrium Conditions

In this section, we describe the conditions determining the equilibrium frequency with which industries are cartelized. Prior to getting into the details, let us provide a brief overview.

1. Taking as given σ (the per period probability that a cartel pays penalties through non-leniency enforcement), we first solve for equilibrium collusive behavior for a type-η industry and the maximum value for π whereby collusion is incentive compatible, denoted $\phi^*(\sigma, \eta)$.

2. With the conditions for internal collapse - which occurs when $\pi > \phi^*(\sigma, \eta)$ - and the likelihood of non-leniency enforcement, σ, along with the probability of cartel formation, κ, a Markov process on cartel birth and death is constructed from which is solved the stationary distribution of industries in terms of their cartel status, for each industry type η. By aggregating over all industry types, the equilibrium cartel rate, $C(\sigma)$, is derived, given σ.

3. The next step is to solve for the equilibrium value of σ, denoted σ^*. The probability that the CA's investigation is successful, $p(\lambda L + R)$, depends on

[13] In Harrington (2011a), sufficient conditions are derived for a CA whose objective is to maximize the number of convictions (which is observable) chooses a policy that minimizes the cartel rate.

the mass of leniency cases, L, and the mass of non-leniency cases, R; both L and R depend on σ as they depend on the cartel rate $C(\sigma)$. σ^* is then a fixed point:

$$\sigma^* = qrp\left(\lambda L\left(\sigma^*\right) + R\left(\sigma^*\right)\right)$$

In other words, σ - the probability that firms are caught, prosecuted, and convicted - determines the cartel rate $C(\sigma)$, the cartel rate determines the caseload $\lambda L(\sigma) + R(\sigma)$, and the caseload determines the probability that they are able to get a conviction on a case and thus determines σ. Given σ^*, the equilibrium cartel rate is $C(\sigma^*)$. Section 4 derives some properties of $C(\sigma)$ and proves the existence of σ^*.

4. When r is fixed, the analysis ends with step 3. When r is endogenous, the final step is to solve for the value that minimizes the cartel rate:

$$r^* \in \arg\min_{r \in [0,1]} C\left(\sigma^*(r)\right).$$

In Section 5, results are derived for when r is fixed. In Section 6, r is chosen to minimize the cartel rate.

3.1 Cartel Formation and Collusive Value

A collusive strategy for a type-η industry entails colluding when π is sufficiently low and not colluding otherwise. The logic is as in Rotemberg and Saloner (1986). When π is high, the incentive to deviate is strong because a firm increases current profit by $(\eta - 1)\pi$. At the same time, the future payoff is independent of the current realization of π, given that π is *iid*. Since the payoff to cheating is increasing in π while the future payoff is independent of π, the incentive compatibility of collusion is more problematic when π is higher.

Suppose firms are able to collude for at least some realizations of π, and let W^o and Y^o denote the payoff when the industry is not cartelized and is cartelized, respectively. If not cartelized then, with probability κ, firms have an opportunity to cartelize with resulting payoff Y^o. With probability $1 - \kappa$, firms do not have such an opportunity and continue to compete. In that case, each firm earns current expected profit of $\alpha\mu$ and a future value of W^o. Thus, the payoff when not colluding is defined recursively by:

$$W^o = (1 - \kappa)\left(\alpha\mu + \delta W^o\right) + \kappa Y^o. \tag{1}$$

As it'll be easier to work with re-scaled payoffs, define:

$$W \equiv (1 - \delta) W^o, \; Y \equiv (1 - \delta) Y^o.$$

Multiplying both sides of (1) by $1 - \delta$ and re-arranging yields:

$$W = \frac{(1 - \kappa)(1 - \delta)\alpha\mu + \kappa Y}{1 - \delta(1 - \kappa)}$$

Also note that the incremental value to being in the cartelized state is:

$$Y - W = Y - \left(\frac{(1-\kappa)(1-\delta)\alpha\mu - \kappa Y}{1-\delta(1-\kappa)} \right) = \frac{(1-\kappa)(1-\delta)(Y-\alpha\mu)}{1-\delta(1-\kappa)}. \tag{2}$$

Suppose firms are cartelized and π is realized. When a firm decides whether to collude or cheat, it decides at the same time whether to apply for leniency. If it decides to collude, it is clearly not optimal to apply for leniency since the cartel is going to be shut down by the authorities in which case the firm ought to maximize current profit by cheating. The more relevant issue is whether it should apply for leniency if it decides to cheat. The incentive compatibility constraint (ICC) is:

$$\begin{aligned} & (1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W] - (1-\delta)\sigma\gamma(Y-\alpha\mu) \\ \geq \; & (1-\delta)\eta\pi + \delta W - (1-\delta)\min\{\sigma,\theta\}\gamma(Y-\alpha\mu). \end{aligned} \tag{3}$$

Examining the LHS expression, if it colludes then it earns current profit of π (given all other firms are colluding). With probability $1-\sigma$, the cartel is not shut down by the CA and, given the industry is in the cartel state, the future payoff is Y. With probability σ, the cartel is caught and convicted by the CA - which means a one-time penalty of $\gamma(Y-\alpha\mu)$ - and since the industry is no longer cartelized, the future payoff is W. Turning to the RHS expression, the current profit from cheating is $\eta\pi$. Since this defection causes the cartel to collapse, the future payoff is W. There is still a chance of being caught and convicted and a deviating firm will apply for leniency iff the penalty from doing so is less than the expected penalty from not doing so (and recall that the other firms are colluding and thus do not apply for leniency): $\theta\gamma(Y-\alpha\mu) < \sigma\gamma(Y-\alpha\mu)$ or $\theta < \sigma$. Given optimal use of the leniency program, the deviating firm's expected penalty is then $\min\{\sigma,\theta\}\gamma(Y-\alpha\mu)$. Re-arranging (3) and using (2), the ICC can be presented as:

$$\begin{aligned} \pi \; & \leq \; \frac{\delta(1-\sigma)(1-\kappa)(Y-\alpha\mu) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma,\theta\}]\gamma(Y-\alpha\mu)}{(\eta-1)[1-\delta(1-\kappa)]} \\ & \equiv \; \phi(Y,\sigma,\eta). \end{aligned} \tag{4}$$

Collusion is incentive compatible iff the current market condition is sufficiently low.[14]

[14] As specified in the ICC in (3), the penalty is slightly different from that in Harrington and Chang (2009) or HC09. In terms of rescaled payoffs, HC09 assumes the penalty is $\gamma(Y-\alpha\mu)$, while here it is $(1-\delta)\gamma(Y-\alpha\mu)$. This means that HC09 assumes that a conviction results in an infinite stream of single-period penalties of $\gamma(Y-\alpha\mu)$ which has a present value of $\gamma\left(\frac{Y-\alpha\mu}{1-\delta}\right)$, while the current paper assumes a one-time penalty of $(1-\delta)\gamma(Y-\alpha\mu)$ which has a present value of $\gamma(Y-\alpha\mu)$. We now believe the latter specification is more sound. For the specification in HC09, every time a cartel is convicted, it has to pay a penalty of $\gamma(Y-\alpha\mu)$ *ad infinitum*. Thus, if it has been convicted k times in the past then it is paying $k\gamma(Y-\alpha\mu)$ in each period, while earning an average collusive profit of μ in each period. As $k \to \infty$, the penalty is unbounded while the payoff from collusion is not. It can be shown that the penalty specification in HC09 implies $\lim_{\delta\to 1} Y = \alpha\mu$ so that the penalty wipes out all gains from colluding. These properties do not seem desirable, and we believe it is better to assume the penalty is a one-time payment $\gamma(Y-\alpha\mu)$ rather than an infinite stream of $\gamma(Y-\alpha\mu)$. It is important to note that this change in specification does not affect the conclusions

In deriving an expression for the value to colluding, we need to discuss usage of the leniency program in equilibrium. Firms do not use it when market conditions result in the cartel being stable but may use it when the cartel collapses. As the continuation payoff is W regardless of whether leniency is used, a firm applies for leniency iff it reduces the expected penalty. First note that an equilibrium either has no firms applying for leniency or all firms doing so because if at least one firm applies then another firm can lower its expected penalty from F to ωF by also doing so. This has the implication that it is always an equilibrium for all firms to apply for leniency. Furthermore, it is the unique equilibrium when $\theta < \sigma$. To see why, suppose all firms were not to apply for leniency. A firm would then lower its penalty from σF to θF by applying. When instead $\sigma \leq \theta$, there is also an equilibrium in which no firm goes for leniency as to do so would increase its expected penalty from σF to θF. Using the selection criterion of Pareto dominance, we will assume that, upon internal collapse of the cartel, no firms apply when $\sigma \leq \theta$ and all firms apply when $\theta < \sigma$.

The expected payoff to being cartelized, $\psi(Y, \sigma, \eta)$, is then recursively defined by:

$$\psi(Y,\sigma,\eta) = \begin{cases} \int_{\underline{\pi}}^{\phi(Y,\sigma,\eta)} \{(1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W] - (1-\delta)\sigma\gamma(Y-\alpha\mu)\} h(\pi)\, d\pi & \text{if } \sigma \leq \theta \\ + \int_{\phi(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W - (1-\delta)\sigma\gamma(Y-\alpha\mu)] h(\pi)\, d\pi & \\ & \\ \int_{\underline{\pi}}^{\phi(Y,\sigma,\eta)} \{(1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W] - (1-\delta)\sigma\gamma(Y-\alpha\mu)\} h(\pi)\, d\pi & \text{if } \theta < \sigma \\ + \int_{\phi(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W - (1-\delta)\omega\gamma(Y-\alpha\mu)] h(\pi)\, d\pi & \end{cases}.$$

To understand this expression, first consider when $\sigma \leq \theta$, in which case leniency is not used. If $\pi \in [\underline{\pi}, \phi(Y,\sigma,\eta)]$ then collusion is incentive compatible; each firm earns current profit of π, incurs an expected penalty of $\sigma\gamma(Y-\alpha\mu)$, and has an expected future payoff of $(1-\sigma)Y + \sigma W$. If instead $\pi \in (\phi(Y,\sigma,\eta), \overline{\pi}]$ then collusion is not incentive compatible, so each firm earns current profit of $\alpha\pi$, incurs an expected penalty of $\sigma\gamma(Y-\alpha\mu)$, and has an expected future payoff of W. The expression when $\theta < \sigma$ differs only when collusion breaks down in which case all firms apply for leniency.[15]

---

in HC09 because of the parameter γ. Starting with the original specification $\gamma(Y-\alpha\mu)$ and defining $\widetilde{\gamma} \equiv \gamma/(1-\delta)$, the analysis in HC09 is equivalent to when the penalty is $(1-\delta)\widetilde{\gamma}(Y-\alpha\mu)$. This transformation works as long as δ is fixed. As the main results in HC09 do not involve performing comparative statics with respect to δ or letting $\delta \to 1$, the conclusions in HC09 remain intact.

[15] Note that if market conditions are sufficiently strong - that is, $\pi > \phi(Y,\sigma,\eta)$ - firms not only do not collude (as it is not incentive compatible) but the cartel breaks down, as reflected in firms having a future payoff of W (less expected penalties) An alternative strategy is to have firms not collude when market conditions are strong but for the cartel to remain in place so that firms collude again as soon as market conditions return to lower levels, in which case the future payoff is Y. The latter equilibrium is more in the spirit of the traditional approach to modelling collusive behavior in that the degree of collusion adjusts to market conditions rather than cartel breakdown occurring. We do not characterize such an equilibrium because, in practice, cartels *do* breakdown - it is not simply that firms go to a coordinated punishment - and it is that death process we want our model to generate. In a richer model in which firms could choose from an array of prices, we would be fine with having some adjustment of the collusive price to market conditions - rather than always having cartel breakdown - as long as, under some market conditions, the cartel does collapse.

A fixed point to ψ is an equilibrium value for Y. That is, given an anticipated future collusive value Y, the resulting equilibrium behavior - as represented by $\phi(Y, \sigma, \eta)$ - results in firms colluding for market states such that the value to being in a cartel is Y. We then want to solve:

$$Y^* = \psi(Y^*, \sigma, \eta).$$

As an initial step to exploring the set of fixed points, first note that $\psi(\alpha\mu, \sigma, \eta) = \alpha\mu$. Hence, one fixed point to ψ is the degenerate solution without collusion. If there is a fixed point with collusion - that is, $Y > \alpha\mu$ - then we select the one with the highest value.

Given $Y^*(\sigma, \eta)$, define

$$\phi^*(\sigma, \eta) \equiv \max\{\min\{\phi(Y^*(\sigma, \eta), \sigma, \eta), \overline{\pi}\}, \underline{\pi}\},$$

as the maximum profit realization such that a type-η cartel is stable. It is a measure of cartel stability since the cartel is stable iff $\pi \leq \phi^*(\sigma, \eta)$ and thus internally collapses with probability $1 - H(\phi^*(\sigma, \eta))$. Note that if $\phi^*(\sigma, \eta) = \overline{\pi}$ then the cartel is stable for all market conditions (so it never internally collapses), and if $\phi^*(\sigma, \eta) = \underline{\pi}$ then the cartel is unstable for all market conditions (so firms never collude).

Before moving on to allowing for a population of industries and endogenizing σ, it is useful to review the various ways in which a leniency program affects the calculus to form and maintain a cartel. Assume σ is fixed and $\theta < \sigma$ so that firms would potentially want to use the leniency program. Previous work has shown that the introduction of a leniency program has three effects (Harrington, 2008). First, it makes cartels less stable by reducing the penalties that a firm receives when it cheats; expected penalties to a deviating firm decline from σF to θF. Referred to as the Deviator Amnesty effect, it tightens the incentive compatibility constraint in (3) and thereby reduces the maximum market state for which collusion is incentive compatible, ϕ^*. Second, the probability of paying penalties is higher because firms in a collapsing cartel will apply for leniency. The probability rises from σ to $\sigma H(\phi^*) + (1 - H(\phi^*))$ where $\sigma H(\phi^*)$ is the probability that a cartel does not collapse but is caught and convicted by the CA and $1 - H(\phi^*)$ is the probability that a cartel internally collapses and firms subsequently apply for leniency. This is the Race to the Courthouse effect and it also tightens the ICC. Third, the Cartel Amnesty effect of a leniency program lowers the penalties that cartel members pay in equilibrium. When the cartel collapses and firms apply for leniency, penalties are reduced from F to ωF. This last effect loosens the ICC and thereby promotes cartel formation. When σ is fixed, Harrington (2008) shows that, generally, these three counter-acting effects net out so that a leniency program makes collusion more difficult. Of course, we are endogenizing σ in the current model by assuming it is lower when the CA's caseload is bigger. This introduces a potentially significant feedback effect that could either reinforce the cartel-reducing effect of a leniency program or counteract it.

3.2 Stationary Distribution of Cartels

Given $\phi^*(\sigma, \eta)$, the stochastic process by which cartels are born and die (either through internal collapse or being shut down by the CA) is characterized in this section. The random events driving this process are the opportunity to cartelize, market conditions, and conviction by the CA. We initially characterize the stationary distribution for type-η industries. The stationary distribution for the entire population of industries is then derived by integrating the type specific distributions over all types.

Consider an arbitrary type-η industry. If it is not cartelized at the end of the preceding period then, by the analysis in Section 3.1, it'll be cartelized at the end of the current period with probability $\kappa(1-\sigma)H(\phi^*(\sigma,\eta))$. With probability κ it has the opportunity to cartelize, with probability $H(\phi^*(\sigma,\eta))$ the realization of π is such that collusion is incentive compatible, and with probability $1-\sigma$ it is not caught and convicted by the CA. If instead the industry was cartelized at the end of the previous period, it'll still be cartelized at the end of this period with probability $(1-\sigma)H(\phi^*(\sigma,\eta))$. Suppose there is a continuum of type-η industries with independent realizations of the stochastic events each period. The task is to characterize the stationary distribution with regards to the frequency of cartels.

Let $NC(\sigma,\eta)$ denote the proportion of type-η industries that are not cartelized. The stationary rate of non-cartels is defined by :

$$\begin{aligned} NC(\sigma,\eta) &= NC(\sigma,\eta)\left[(1-\kappa)+\kappa(1-H(\phi^*))+\kappa\sigma H(\phi^*)\right] \\ &\quad +\left[1-NC(\sigma,\eta)\right]\left[(1-H(\phi^*))+\sigma H(\phi^*)\right] \end{aligned} \tag{5}$$

Examining the RHS of (5), a fraction $NC(\sigma,\eta)$ of type-η industries were not cartelized in the previous period. Out of those industries, a fraction $1-\kappa$ will not have the opportunity to cartelize in the current period. A fraction $\kappa(1-H(\phi^*))$ will have the opportunity but, due to a high realization of π, find it is not incentive compatible to collude, while a fraction $\kappa\sigma H(\phi^*)$ will cartelize and collude but then are discovered by the CA. Of the industries that were colluding in the previous period, which have mass $1-NC(\eta)$, a fraction $1-H(\phi^*)$ will collapse for internal reasons and a fraction $\sigma H(\phi^*)$ will instead be caught by the authorities and thus shut down.

Solving (5) for $NC(\sigma,\eta)$:

$$NC(\sigma,\eta)=\frac{1-(1-\sigma)H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi^*(\sigma,\eta))}. \tag{6}$$

For the stationary distribution, the fraction of cartels among type-η industries is then:

$$C(\sigma,\eta)\equiv 1-NC(\sigma,\eta)=\frac{\kappa(1-\sigma)H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi^*(\sigma,\eta))}. \tag{7}$$

Finally, the derivation of the entire population of industries is performed by integrating the type-η distribution over $\eta\in\left[\underline{\eta},\overline{\eta}\right]$. The mass of cartelized industries, which we refer to as the cartel rate $C(\sigma)$, is then defined by:

$$C(\sigma)\equiv\int_{\underline{\eta}}^{\overline{\eta}}C(\sigma,\eta)g(\eta)d\eta=\int_{\underline{\eta}}^{\overline{\eta}}\left[\frac{\kappa(1-\sigma)H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi^*(\sigma,\eta))}\right]g(\eta)d\eta. \tag{8}$$

3.3 Equilibrium Probability of Paying Penalties

Recall that $\sigma = qrs$ where q is the probability of a cartel being discovered, r is the probability that the CA investigates a reported case, and s is the probability of it succeeding with the investigation. We now want to derive the equilibrium value of s, where $s = p\left(\lambda L + R\right)$, L is the mass of leniency cases, and R is the mass of non-leniency cases handled by the CA. As both L and R depend on the cartel rate C and the cartel rate depends on s (through σ), this is a fixed point problem. We need to find a value for s, call it s', such that, given $\sigma = qrs'$, the induced cartel rate $C\left(qrs'\right)$ is such that it generates L and R so that $p\left(\lambda L + R\right) = s'$.

With our expression for the cartel rate, we can provide expressions for L and R. The mass of cartel cases generated by the leniency program is:

$$L\left(\sigma\right) = \begin{cases} 0 & \text{if } \sigma \leq \theta \\ \\ \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi^*\left(\sigma, \eta\right)\right)\right) C\left(\sigma, \eta\right) g\left(\eta\right) d\eta & \text{if } \theta < \sigma \end{cases} \tag{9}$$

In (9), note that an industry does not apply for leniency when it is still effectively colluding. When collusion stops, leniency is used when the only equilibrium is that all firms apply for leniency, which is the case when $\theta < \sigma$. Thus, when $\theta < \sigma$, L equals the mass of cartels that collapse due to a high realization of π. This is consistent with a concern expressed by a European Commission official that many leniency applicants are from dying cartels.[16, 17]

The mass of cartel cases generated without use of the leniency program is

$$R\left(\sigma\right) = \begin{cases} qrC\left(\sigma\right) & \text{if } \sigma \leq \theta \\ \\ qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*\left(\sigma, \eta\right)\right) C\left(\sigma, \eta\right) g\left(\eta\right) d\eta & \text{if } \theta < \sigma \end{cases} \tag{10}$$

If the leniency program is never used (which is when $\sigma \leq \theta$), then the mass of cases being handled by the CA is $qrC\left(\sigma\right)$. If instead $\theta < \sigma$, so that dying cartels use the leniency program, then the cartels left to be caught are those which have not collapsed in the current period which is $\int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*\left(\sigma, \eta\right)\right) C\left(\sigma, \eta\right) g\left(\eta\right) d\eta$.

The equilibrium probability of a CA successfully getting a cartel to pay penalties (without use of the leniency program) is the solution to the following fixed point problem:

$$\sigma = \Psi\left(\sigma\right) \equiv \begin{cases} qrp\left(qrC\left(\sigma\right)\right) & \text{if } \sigma \leq \theta \\ \\ qrp\Big(\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi^*\left(\sigma, \eta\right)\right)\right) C\left(\sigma, \eta\right) g\left(\eta\right) d\eta & \text{if } \theta < \sigma \\ \quad + qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*\left(\sigma, \eta\right)\right) C\left(\sigma, \eta\right) g\left(\eta\right) d\eta\Big) & \end{cases} \tag{11}$$

[16] This statement was made by Olivier Guersent at the 11th Annual EU Competition Law and Policy Workshop: Enforcement of Prohibition of Cartels in Florence, Italy in June 2006.

[17] That either all firms or no firms apply for leniency is a property of not only our analysis but all previous analyses on leniency programs with the exception of some recent work by one of the authors, Harrington (2011b, 2012). In those two papers, there is private information between cartel members which can explain why only one firm would come forward to the CA.

where we have substituted for L using (9) and R using (10).[18] If there are multiple solutions to (11) then it is assumed the maximal one is selected.[19]

3.4 Optimal Competition Policy

The results in Section 5 are derived taking enforcement policy - as parameterized by r which is the fraction of possible cases that the CA takes on - as fixed, while the results in Section 6 assume the CA acts to minimize the cartel rate. Let $\sigma^*(r)$ denote the maximal solution to (11), where its dependence on r is now made explicit. For when r is endogenized, it is assumed that $r = r^*$ where

$$r^* \in \arg \min_{r \in [0,1]} C(\sigma^*(r)).$$

By having the prosecution policy chosen to minimize the cartel rate, the analysis delivers an upper bound on welfare.

3.5 Summary of Solution Algorithm

1. Given σ and Y and for each η, solve for the maximum market condition (or threshold) for which the ICC is satisfied, $\phi(Y, \sigma, \eta)$.

2. Given σ and for each η, solve for the equilibrium collusive value $Y^*(\sigma, \eta)$ which is a solution to the fixed point problem:

$$Y^* = \psi(Y^*, \sigma, \eta).$$

   If there are multiple fixed points, select the maximum. Given $Y^*(\sigma, \eta)$, define the equilibrium threshold:

$$\phi^*(\sigma, \eta) \equiv \phi(Y^*(\sigma, \eta), \sigma, \eta).$$

3. Given σ and $\phi^*(\sigma, \eta)$, derive the stationary proportion of type-η industries that are cartels:

$$C(\sigma, \eta) = \frac{\kappa(1-\sigma) H(\phi^*(\sigma, \eta))}{1 - (1-\kappa)(1-\sigma) H(\phi^*(\sigma, \eta))},$$

   and integrate over industry types to derive the stationary cartel rate:

$$C(\sigma) = \int_{\underline{\eta}}^{\overline{\eta}} C(\sigma, \eta) g(\eta) d\eta.$$

[18] Note that the fixed point can be defined in terms of either σ or s given that $\sigma = qrs$ and, at this point of the analysis, q and r are parameters.

[19] We conjecture that results hold with some other selections, such as the minimal fixed point to Ψ. What is necessary is that a shift up (down) in Ψ increases (decreases) σ^*.

4. Solve for the equilibrium probability of paying penalties through non-leniency enforcement σ^* which is a solution to the fixed point problem:

$$\sigma^* = \Psi(\sigma^*) .$$

If there are multiple fixed points, select the maximum. The equilibrium cartel rate is $C(\sigma^*)$.

5. (Optional) Solve for the optimal prosecution rate r^* :

$$r^* \in \arg\min_{r \in [0,1]} C(\sigma^*(r)) .$$

4 Equilibrium Cartel Rate: Existence and Properties

Section 4.1 derives some properties of the cartel rate $C(\sigma)$ function. In particular, it is shown that $C(\sigma)$ is decreasing in σ. Taking σ as exogenous, if firms assign a higher probability to the CA discovering, prosecuting, and convicting cartels then a smaller fraction of industries are cartelized, either because fewer cartels form and/or those cartels that form have shorter average duration.[20] In Section 4.2, it is shown that a solution to (11) exists for when there is no leniency program ($\theta = 1$) and there is a full leniency program ($\theta = 0$). While these results are of intrinsic interest, their primary purpose is to provide the foundation for the analysis in Sections 5 and 6 which explores the impact of a leniency program.

4.1 Properties of the Cartel Rate Function

The main result of this sub-section is that the cartel rate is decreasing in the probability that cartels assign to paying penalties through non-leniency enforcement, σ. Before launching into the analysis, let us provide an overview. Recall that the value to being in the cartel state is a fixed point: $Y^* = \psi(Y^*)$. Lemma 1 shows that ψ maps $[\alpha\mu, \mu]$ into itself and, given that ψ is a continuous function of Y, a fixed point exists which is the equilibrium collusive value. Recall that $\phi(Y, \sigma, \eta)$ is the maximal market condition whereby collusion is stable; that is, the ICC is satisfied iff $\pi \leq \phi(Y, \sigma, \eta)$. $\phi^*(\sigma, \eta) = \phi(Y^*, \sigma, \eta)$ is the equilibrium threshold after substituting in the equilibrium collusive value. Lemma 2 show that Y^* and ϕ^* are decreasing in η and σ so that more intense non-leniency enforcement (that is, a higher probability of paying penalties) lowers the collusive value and makes a cartel less stable (as the cartel collapses with lower market conditions). Lemma 3 shows that the maximal industry type for which a successful cartel can form, $\widehat{\eta}$, is lower when σ is higher. ($\widehat{\eta}$ is that value such that $\phi^*(\sigma, \eta) > \underline{\pi}$ iff $\eta \leq \widehat{\eta}$.) It is then proven using Lemmas 2 and 3 that $C(\sigma)$ is decreasing in σ (Lemma 4).

[20] The results in Section 4.1 correspond to some properties proven in Theorems 1, 3, 4, and 6 of Harrington and Chang (2009). However, the results in Harrington and Chang (2009) assume σ is sufficiently small which we do not want to do here given that σ is now endogenized. Instead, results are derived for when the penalty multiple γ is sufficiently small.

Results are derived for when the penalty multiple γ is not too high, which must be the case if collusion is to emerge in equilibrium. Lemma 1 considers properties of the collusive value function $\psi(Y)$. Given that the penalty if convicted is $\gamma(Y - \alpha\mu)$ and thus is proportional to the collusive value, if γ is too high then $\psi(Y)$ will have the pathological property that it is *decreasing* in Y; that is, a higher future collusive value Y actually reduces the value to being in the cartel state because it raises the penalty even more. In that case, collusion will not occur. γ must then be sufficiently low so that ψ is increasing in Y and thus there can exist a fixed point exceeding $\alpha\mu$ which means that firms collude with positive probability. All proofs are in the Appendix.

Lemma 1 *$\exists\widehat{\gamma} > 0$ such that if $\gamma \in [0, \widehat{\gamma})$ then: (i) $\psi : [\alpha\mu, \mu] \rightarrow [\alpha\mu, \mu]$; and (ii) $\psi'(Y) > 0$, for all $Y \in [\alpha\mu, \mu]$.*

By Lemma 1, $Y^*(\sigma, \eta)$ exists which is the recursively defined value to a firm in a type-η industry when it is in the cartel state given the probability of paying penalties through non-leniency enforcement is σ. Lemma 2 shows that this collusive value is higher when non-leniency enforcement is weaker (that is, σ is lower) and the profit gain to cheating is less (η is lower). Recall that collusion internally collapses if and only if $\pi > \phi^*(\sigma, \eta)$. Lemma 2 also shows that a cartel is more stable - in the sense that $\phi^*(\sigma, \eta)$ is higher so it takes more extreme market conditions to violate the ICC - when σ and η are lower.

Lemma 2 *$\exists\widehat{\gamma} > 0$ such that if $\gamma \in [0, \widehat{\gamma})$ then: i) $Y^*(\sigma, \eta)$ and $\phi^*(\sigma, \eta)$ are non-increasing in σ and η; ii) if $Y^*(\sigma, \eta) > \alpha\mu$ then $Y^*(\sigma, \eta)$ is decreasing in σ; and iii) if $\phi^*(\sigma, \eta) \in (\underline{\pi}, \overline{\pi})$ then $Y^*(\sigma, \eta)$ is decreasing in η and $\phi^*(\sigma, \eta)$ is decreasing in σ and η.*

Given that $\phi^*(\sigma, \eta)$ is decreasing in η when $\phi^*(\sigma, \eta) \in (\underline{\pi}, \overline{\pi})$ then, if $\phi^*(\sigma, \overline{\eta}) = \underline{\pi}$, there exists $\widehat{\eta}(\sigma) \in [\underline{\eta}, \overline{\eta})$ such that $\phi^*(\sigma, \eta) > \underline{\pi}$ iff $\eta \leq \widehat{\eta}(\sigma)$. Hence, if $\eta \leq \widehat{\eta}(\sigma)$ then an industry can successfully collude with some probability (that is, for some market conditions) and if $\eta > \widehat{\eta}(\sigma)$ then an industry is never able to successfully collude. The next lemma shows that the set of industry types that can successfully collude, $[\underline{\eta}, \widehat{\eta}(\sigma)]$, shrinks as non-leniency enforcement intensifies (σ is increased).

Lemma 3 *$\exists\widehat{\gamma} > 0$ such that if $\gamma \in [0, \widehat{\gamma})$ then $\widehat{\eta}(\sigma)$ is decreasing in σ.*

Lemma 4 shows that more intense non-leniency enforcement reduces the cartel rate. The decline in the cartel rate comes from those cartels that form having shorter duration - because $\phi^*(\sigma, \eta)$ declines (by Lemma 2) - and that fewer industries cartelize - because $\widehat{\eta}(\sigma)$ declines (by Lemma 3).

Lemma 4 *$\exists\widehat{\gamma} > 0$ such that if $\gamma \in [0, \widehat{\gamma})$ then $C(\sigma)$ is non-increasing in σ and if $C(\sigma) > 0$ then $C(\sigma)$ is decreasing in σ.*

4.2 Existence of an Equilibrium Cartel Rate

From Section 3.3, the equilibrium level of non-leniency enforcement σ^* (which is the probability of a cartel being caught, prosecuted, and convicted by a CA) is the fixed point to (11) which is repeated here:

$$\sigma = \Psi(\sigma) \equiv \begin{cases} qrp\left(qr\int_{\underline{\eta}}^{\overline{\eta}} C(\sigma,\eta)\, g(\eta)\, d\eta\right) & \text{if } \sigma \leq \theta \\ \\ qrp\left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} (1 - H(\phi^*(\sigma,\eta)))\, C(\sigma,\eta)\, g(\eta)\, d\eta \right. & \text{if } \theta \leq \sigma \\ \left. +qr\int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma,\eta))\, C(\sigma,\eta)\, g(\eta)\, d\eta\right) & \end{cases}$$

where

$$C(\sigma,\eta) = \frac{\kappa(1-\sigma)\, H(\phi^*(\sigma,\eta))}{1 - (1-\kappa)(1-\sigma)\, H(\phi^*(\sigma,\eta))}.$$

A fixed point σ^* has the property that if firms believe that the per period probability of paying penalties (through non-leniency enforcement) is σ^* then the induced cartel birth and death rates generate a caseload for the CA whereby the equilibrium conviction rate s^* satisfies $qrs^* = \sigma^*$.

While Ψ maps $[0,1]$ into itself, the existence of σ^* is not immediate due to two possible sources of discontinuity in Ψ. Recall that $\phi^*(\sigma,\eta) = \phi(Y^*(\sigma,\eta), \sigma, \eta)$ and $Y^*(\sigma,\eta)$ is the maximal fixed point to: $Y = \psi(Y,\sigma,\eta)$. Because of multiple fixed points to $\psi(Y,\sigma,\eta)$, $Y^*(\sigma,\eta)$ need not be continuous in σ and if $Y^*(\sigma,\eta)$ is discontinuous then $\phi^*(\sigma,\eta)$ is discontinuous which implies $H(\phi^*(\sigma,\eta))$ and $C(\sigma,\eta)$ are discontinuous. It is proven in Theorem 5 that possible discontinuities in the integrand of Ψ do not create discontinuities in Ψ. There is a second possible source of discontinuity in Ψ which is due to a discontinuity in expected penalties at $\sigma = \theta$. That discontinuity is present as long as $\theta \in (0,1)$ and, as a result, existence is proven only when there is no leniency ($\theta = 1$) and full leniency ($\theta = 0$).

Theorem 5 *For $\theta \in \{0,1\}$, $\exists\widehat{\gamma} > 0$ such that if $\gamma \in [0,\widehat{\gamma})$ then σ^* exists.*

In the ensuing analysis, it is assumed (without being stated) that γ is sufficiently low so that the results of Section 4 apply.

5 Impact of a Leniency Program: Analytical Results

In this section, we begin to explore the impact of introducing a corporate leniency program. Does a leniency program always promote desistance (encouraging cartels to shut down) and deterrence (discouraging cartels from forming) or can it be counterproductive and actually result in more cartels? If it can, what are the conditions that avoid such dysfunctional implications and instead ensure that a leniency program reduces the frequency of cartels? What policies can a CA pursue to promote such an outcome?

The analysis will focus on comparing the cases of no leniency program ($\theta = 1$) with a leniency program in which the first firm to come forward receives full leniency ($\theta = 0$). There is no reason to think that results do not extend to when leniency is almost full ($\theta \simeq 0$) though existence of equilibrium has not been established (see the discussion in Section 4.2). To economize on notation and make it easier for the reader to follow the analysis, expressions with an NL subscript will refer to the case of "no leniency program," while those with an L subscript will refer to the case of a "full leniency program." For example, $C_{NL}(\sigma)$ and $C_L(\sigma)$ are, respectively, the cartel rate functions without leniency and with (full) leniency. The associated fixed points for σ are given by:

$$\begin{aligned}
\sigma_{NL}^* &= qrp\left(qr\int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta\right) \\
\sigma_L^* &= qrp\left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right)\right) C_L\left(\sigma_L^*, \eta\right) g\left(\eta\right) d\eta + qr\int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right) C_L\left(\sigma_L^*, \eta\right) g\left(\eta\right) d\eta\right),
\end{aligned}$$

and the equilibrium cartel rates are $C_{NL}(\sigma_{NL}^*)$ and $C_L(\sigma_L^*)$. The analysis will also be conducted holding fixed the CA's non-leniency enforcement instrument r, which is the proportion of discovered cases that it prosecutes. Note that if it is shown that a leniency program decreases (increases) the cartel rate for all values of r then allowing r to be chosen to minimize the cartel rate will still result in a leniency program decreasing (increasing) the cartel rate. Thus, the conclusions of this section are applicable to when a CA acts in a welfare-maximizing manner by choosing its prosecutorial caseload to minimize the cartel rate. As explained earlier, we are not claiming that a welfare-maximizing CA is a good behavioral description but rather that it provides a useful benchmark in terms of what a leniency policy can deliver.

Before taking account of how a leniency program endogenously influences non-leniency enforcement (as measured by σ), Section 5.1 shows, under rather general conditions, that a leniency program lowers the cartel rate, holding non-leniency enforcement fixed. This result is of intrinsic interest but is also instrumental for the ensuing analysis. Non-leniency enforcement is then endogenized and, in Section 5.2, sufficient conditions are provided for a leniency program to lower the cartel rate and, in Section 5.3, for a leniency program to raise the cartel rate. Utilizing those results, Section 5.4 draws some general intuition regarding what is required for a leniency program to serve its intended goal of reducing the cartel rate. The analysis of the impact of a leniency program continues in Section 6 where numerical results are reported.

5.1 Impact of a Leniency Program on the Cartel Rate Function

We begin by analyzing how the cartel rate responds to a leniency program while making the standard assumption in the literature that non-leniency enforcement is exogenous and fixed. Theorem 6 shows that if the probability of paying penalties through non-leniency enforcement is neither too low nor too high - $\sigma \in (\theta, \omega)$ - then a leniency program does not raise the cartel rate. (Recall that a firm pays a fraction θ of the standard penalty when it receives leniency and pays, in expectation, a fraction

ω when all firms apply for leniency.) Furthermore, a leniency program reduces the cartel rate if it is further assumed that there is positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) < \overline{\pi}$ (that is, without a leniency program, some industry types cannot fully collude) and a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) > \underline{\pi}$ (that is, some industry types can collude). These conditions ensure that the cartel rate is positive but not maximal, and rule out cases in which a leniency program does not lower the cartel rate because the cartel rate is either zero without a leniency program or the environment is so conducive to collusion that the cartel rate is maximal with or without a leniency program.

To appreciate the condition $\sigma \in (\theta,\omega)$, first note that if $\sigma > \theta$ then a firm that contemplates deviating from a cartel would apply for leniency in that instance, as doing so reduces its expected penalty from $\sigma\gamma(Y-\alpha\mu)$ to $\theta\gamma(Y-\alpha\mu)$. $\sigma > \theta$ also has the implication that, in response to the internal collapse of a cartel, all firms apply for leniency because it is not an equilibrium for all firms not to apply. In that case, if $\sigma < \omega$ then a firm's expected penalty rises with a leniency program from $\sigma\gamma(Y-\alpha\mu)$ to $\omega\gamma(Y-\alpha\mu)$. Thus, if $\sigma \in (\theta,\omega)$ then a firm will use the leniency program if it deviates or if the cartel collapses and, in the latter situation, expected penalties are higher compared to when there is no leniency program.

The result of Section 5.1 is derived for the general case when leniency may be partial. For when there is a leniency program, variables have subscript θ to indicate that the policy is that the leniency recipient has a fraction $1-\theta$ of penalties waived; for example, $C_\theta(\sigma)$ is the cartel rate function in that case.

Theorem 6 *If $\sigma \in (\theta,\omega)$ then $C_{NL}(\sigma) \geq C_\theta(\sigma)$ and if there is positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) > \underline{\pi}$ then $C_{NL}(\sigma) > C_\theta(\sigma)$.*

Under fairly general conditions, Theorem 6 shows that a leniency program reduces the frequency of cartels when non-leniency enforcement is held fixed. Let us summarize the forces that lie behind that result. Consider the matter from an individual cartel of type η. It was shown in the proof of Theorem 6 that $\phi_{NL}(Y,\sigma,\eta) > \phi_\theta(Y,\sigma,\eta)$ which comes from the Deviator Amnesty effect. Holding fixed the value in the cartel state Y, the payoff to a firm from choosing to collude rather than compete is unchanged by a leniency program. However, a leniency program increases the payoff to cheating on the cartel as now a firm can reduce its penalty by applying for leniency. This tightens the ICC and (weakly) shrinks the set of market conditions for which collusion is stable from $\pi \in [\underline{\pi}, \phi_{NL}(Y,\sigma,\eta)]$ to $\pi \in [\underline{\pi}, \phi_\theta(Y,\sigma,\eta)]$ which then reduces expected cartel duration and the collusive value function. There is a second effect to a leniency program on the collusive value function which comes when the cartel collapses. If $\theta < \sigma$ then the only equilibrium is for firms to all apply for leniency in which case expected penalties go from $\sigma\gamma(Y-\alpha\mu)$ to $\omega\gamma(Y-\alpha\mu)$. If $\omega > \sigma$ then a leniency program increases expected penalties and this decreases the collusive value function. Hence, a leniency program reduces the collusive value function - $\psi_{NL}(Y,\sigma,\eta) > \psi_\theta(Y,\sigma,\eta)$ - both because it reduces cartel duration and raises penalties in the event of cartel collapse. As a result, equilibrium collusive value is lower -

$Y^*_{NL}(\sigma,\eta) > Y^*_\theta(\sigma,\eta)$ - and the equilibrium threshold is lower - $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta)$. Hence, either a cartel no longer forms - when $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta) = \underline{\pi}$ - or has shorter duration - when $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta) > \underline{\pi}$ and, therefore, the cartel rate is lower.

Taking non-leniency enforcement as exogenous, Theorem 6 tells says that a leniency program can generally be expected to reduce the cartel rate. This finding is consistent with that found in the previous theoretical literature. In the remainder of Section 5, we explore what happens when non-leniency enforcement is endogenous and reacts to the introduction of a leniency program. Non-leniency enforcement could either reinforce a leniency program - when non-leniency enforcement is strengthened by a leniency program - or work against it - when non-leniency enforcement is weakened because of a leniency program.

5.2 Leniency Program Decreases the Cartel Rate

Recall that an industry has a realization of market condition π in each period where π is a firm's profit when all firms collude, $\alpha\pi$ is a firm's profit when all firms compete, and $\eta\pi$ is a firm's profit when it competes (or cheats) and all rivals collude. A higher value of π makes it more difficult to satisfy the ICC and collusion is stable if and only if π is sufficiently small. π has cdf H with support $[\underline{\pi}, \overline{\pi}]$. Our first result assumes H is a uniform distribution which is shown to have the key implication that a cartel is either stable for all $\pi \in [\underline{\pi}, \overline{\pi}]$ or is unstable for all $\pi \in [\underline{\pi}, \overline{\pi}]$. Thus, cartels never internally collapse which means that a firm never applies for leniency (in equilibrium). Though the leniency program is inactive as measured by the number of applicants, Theorem 7 shows that a leniency program is assured of reducing the cartel rate even when non-leniency enforcement is endogenous.

Theorem 7 *If H is the uniform distribution and the equilibrium cartel rate without a leniency program is positive - $C_{NL}(\sigma^*_{NL}) > 0$ - then a full leniency program reduces the cartel rate: $C_L(\sigma^*_L) < C_{NL}(\sigma^*_{NL})$.*

Under the assumption that market conditions are uniformly distributed, the proof of Theorem 7 shows that a cartel is either fully stable or not stable at all. Hence, cartels never collapse, they terminate only by being discovered and convicted by the CA. Given that only collapsed cartels apply for leniency, there are then no leniency applications. Nevertheless, the leniency program promotes cartel deterrence and thereby reduces the cartel rate. The presence of a leniency program enhances the payoff to a firm from cheating because it can set a low price, earn high profit, *and* avoid penalties by going for leniency; this is the Deviator Amnesty Effect. (Note that there are no Cartel Amnesty and Race to the Courthouse Effects because cartels never use the leniency program.) As a result, the ICC tightens so that some industries now are no longer able to successfully collude. Holding non-leniency enforcement (that is, σ) fixed, a leniency program then lowers the cartel rate. Allowing non-leniency enforcement to respond reinforces the deterrence effect because the lower cartel rate means fewer non-leniency cases (and there are no leniency cases to add to the caseload) which raises the probability of gaining a conviction and thus raises σ. A higher σ then lowers the collusive value which results in more industries being unable to collude. The endogeniety of non-leniency enforcement to a leniency program then serves to reinforce the efficacy of a leniency program: A leniency program reduces the cartel rate which reduces the CA's caseload which raises non-leniency enforcement which lowers the collusive value which lowers the cartel rate and so forth.

The next result shows that a leniency program reduces the cartel rate when $\sigma_{NL}^* \in (0, \omega)$ and λ is sufficiently low. Recall that if it takes x resources to prosecute a non-leniency case then it takes only λx resources to prosecute a leniency case where $\lambda \in (0, 1]$. If λ is sufficiently low then leniency cases take up few resources compared to non-leniency cases. The condition $\sigma_{NL}^* \in (0, \omega)$ says that non-leniency enforcement (in the absence of a leniency program) is positive (so $\sigma_{NL}^* > \theta = 0$) but sufficiently weak that expected penalties without a leniency program are less than expected penalties with a leniency program when all firms apply for leniency (that is, $\sigma_{NL}^* < \omega$). If all firms applying for leniency gives each an equal chance of receiving it, then $\omega = \frac{n-1+\theta}{n} = \frac{n-1}{n} \geq \frac{1}{2}$. Thus, $\sigma_{NL}^* < \frac{1}{2}$ is sufficient for $\sigma_{NL}^* < \omega$ to be satisfied. For reasonable parameter values, one would expect $\sigma_{NL}^* < \frac{1}{2}$ for, if that is not the case, then it is unlikely that collusion will be profitable.

Theorem 8 *If $\sigma_{NL}^* \in (0, \omega)$ then there exists $\widehat{\lambda} > 0$ such that if $\lambda \in \left[0, \widehat{\lambda}\right]$ then the equilibrium cartel rate is weakly lower with a leniency program. If, in addition, there is a positive measure of values for η such that $\phi_{NL}^*(\sigma_{NL}^*, \eta) < \overline{\pi}$ and a positive measure of values for η such that $\phi_{NL}^*(\sigma_{NL}^*, \eta) > \underline{\pi}$ then the equilibrium cartel rate is strictly lower with a leniency program.*

If there are no leniency applications, a lower cartel rate from having a leniency program means a lower caseload which results in a higher probability of conviction which further reduces the cartel rate. Thus, if the cartel rate is lower given σ then, once σ is endogenized, the equilibrium value of σ is higher and thus the equilibrium cartel rate is even lower. That analysis can change once taking into account leniency applications which contribute to caseload and can reduce σ. However, if λ is sufficiently small, leniency applications do not contribute much to caseload in which case the σ-decreasing force coming from leniency applications adding to caseload is dominated by the σ-increasing force coming from fewer non-leniency cases because of fewer cartels. Thus, $\sigma_L^* > \sigma_{NL}^*$ and, given that $C_{NL}(\sigma) > C_L(\sigma)$, it follows that the equilibrium cartel rate is lower with a leniency program: $C_{NL}(\sigma_{NL}^*) > C_L(\sigma_L^*)$.

The final result of this sub-section considers when non-leniency enforcement is weak because the likelihood of discovering a cartel is small: $q \simeq 0$. In that case, a leniency program is sure to be beneficial. Theorem 6 suggests that a necessary condition for a leniency program to fail to lower the cartel rate is that it adversely affects non-leniency enforcement. However, if non-leniency enforcement is absent prior to the introduction of a leniency program then a leniency program cannot further weaken non-leniency enforcement and thus a leniency program must lower the cartel rate. Hence, if a CA is not actively engage in enforcement, a leniency program is sure to be effective in reducing the frequency of cartels.

Theorem 9 *There exists $\widehat{q} > 0$ such that if $q \in [0, \widehat{q}]$ then the equilibrium cartel rate is weakly lower with a leniency program. If, in addition, there is a positive measure of values for η such that $\phi^*_{NL}(\sigma^*_{NL}, \eta) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}(\sigma^*_{NL}, \eta) > \underline{\pi}$ then the equilibrium cartel rate is strictly lower with a leniency program.*

Theorem 9 is dependent on there being full leniency for the first firm: $\theta = 0$. If $\theta > 0$ then, as $q \to 0$ and $\sigma \to 0$, the leniency program has no effect because a deviator would not use it, and cartel members do not use it upon collapse of the cartel because $\omega > 0$. This comment highlights the complementarity between leniency and non-leniency enforcement; if $\sigma < \theta$ then a leniency program is irrelevant because the chances of being caught through non-leniency means is sufficiently low to make applying for leniency not to be in a firm's interests. Except when leniency is literally full - which is rarely the case - the efficacy of a leniency program depends on cartel members believing there is at least some chance of them being caught and convicted by the CA.[21]

5.3 Leniency Program Increases the Cartel Rate

The next result shows that a leniency program can actually cause there to be more cartels. Sufficient conditions for this to occur are that penalties are sufficiently weak and a leniency case uses up about as much resources as a non-leniency case. A leniency program is counter-productive because leniency cases crowd out non-leniency cases at the CA which reduces desistance. On the other hand, cartels that die are now assured of paying penalties because one of them is an informant through the leniency program. That has the potential to enhance deterrence but the effect is quite small because penalties are weak. In spite of the leniency program apparently "working" in the sense of bringing forth leniency applications, it is actually counter-productive in that the latent cartel rate is higher.[22]

Theorem 10 *Assume*

$$\int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi^*_{NL}\left(\sigma^*, \eta\right)\right)\right) C_{NL}\left(\sigma^*_{NL}, \eta\right) g\left(\eta\right) d\eta > 0 \tag{12}$$

so that, without a leniency program, there are cartels that collude and internally collapse. Generically, there exists $\widehat{\lambda} < 1$ and $\widehat{\gamma} > 0$ such that if $(\gamma, \lambda) \in [0, \widehat{\gamma}] \times \left[\widehat{\lambda}, 1\right]$ then the cartel rate with a leniency program strictly exceeds the cartel rate without a leniency program.

[21] An important caveat here is that we have assumed that firms achieve the Pareto-superior equilibrium when it comes to applying for leniency; that is, if there is an equilibrium in which no firms seek leniency then that is the equilibrium upon which firms coordinate. However, experimental evidence suggests that a leniency program can be effective even when $\sigma = 0$ (Bigoni et al, 2012). In that case, presumably a firm is applying for leniency out of concern that a rival will apply for leniency which is sensible for the rival only if it possesses a similar concern.

[22] Theorem 10 is a generic result because it requires that, in an ε-ball around $\gamma = 0$, $Y^*_{NL}(\sigma, \eta)$ and $Y^*_L(\sigma, \eta)$ are continuous in γ.

If penalties are low then a leniency program can cause the cartel rate to rise because deterrence is not enhanced while desistance is weakened through reduced efficacy in prosecuting non-leniency cases. The introduction of a leniency program causes scarce CA resources to be used on cartels that have already shut down rather than used to convict (and thereby shut down) active cartels that were discovered through non-leniency devices. (12) ensures that, without a leniency program, there are indeed dying cartels so that this crowding out of non-leniency cases does occur. Thus, a leniency program is actually reducing desistance. The potential deterrent role that a leniency program can play here is that it guarantees that dying cartels pay penalties (other than the firm that receives leniency). However, if penalties are low then the rise in deterrence is small compared to the fall in desistance and, as a result, more cartels form and they last longer which contributes to a higher cartel rate.[23]

Central to a leniency program raising the cartel rate is that leniency applications are coming from cartels that have already collapsed and thus their conviction serves no desistance role. This finding may be sensitive to a simplifying assumption. As cartels can reconstitute themselves, the manner of death can matter. If a cartel that internally collapsed but was not convicted is able to reconstitute itself faster on average than a cartel that internally collapsed but was not convicted then the leniency program would promote desistance by raising the expected time until the cartel reforms and that could lower the cartel rate. The model could be enriched to allow for this effect; see footnote 9. If that effect was present, however, it may no longer be the case that it is an equilibrium for firms to apply for leniency when the cartel has collapsed. A firm would have to take into account the lower penalties from gaining leniency with the lower expected future profit from delaying the time until the cartel reforms. If the latter were greater then firms would not apply for leniency upon cartel collapse in which case the leniency program would be useless.

Theorem 10 shows that, for any value of r (which is the fraction of possible cases that the CA chooses to prosecute), the cartel rate is higher without a leniency program when penalties are sufficiently weak and a leniency case uses almost as much resources as a non-leniency case to prosecute:[24]

$$C_L\left(qrs_L^*\left(r\right)\right) > C_{NL}\left(qrs_{NL}^*\left(r\right)\right),\ \forall r. \tag{13}$$

Recall that $\sigma = qrs$ and the dependence of the conviction rate s on r has been made explicit. Now, if the CA chooses its caseload in order to minimize the cartel rate, the optimal enforcement policy with and without a leniency program is:

$$\begin{aligned} r_{NL}^* &\in \arg\min_{r\in[0,1]} C_{NL}\left(qrs_{NL}^*\left(r\right)\right) \\ r_L^* &\in \arg\min_{r\in[0,1]} C_L\left(qrs_L^*\left(r\right)\right). \end{aligned}$$

[23] Note that (12) rules out the case in which H is a uniform distribution; thus, Theorem 9 does not conflict with Theorem 7.

[24] It does require that for any value of r, (12) is satisfied which is not very restrictive because whether a cartel collapses is partly due to forces unrelated to the CA.

It then follows from (13),

$$C_L\left(qr_L^* s_L^*\left(r_L^*\right)\right) > C_{NL}\left(qr_{NL}^* s_{NL}^*\left(r_{NL}^*\right)\right).$$

Thus the cartel rate is higher with a leniency program even with a welfare-maximizing CA.

5.4 Discussion

Thus far, we have shown that: 1) a leniency program can either lower or raise the cartel rate; and 2) the number of leniency applications may not be a proper measure of the performance of a leniency program because a leniency program can be inactive yet the cartel rate is lower (Theorem 7) or a leniency program can be active yet the cartel rate is higher (Theorem 10). In concluding Section 5, let us provide a general discussion within which to place the preceding results and which has the objective of identifying the conditions under which a leniency program can be expected to be effective in reducing the cartel rate.

Under fairly general conditions, Theorem 6 showed that a leniency program lowers the cartel rate when holding fixed non-leniency enforcement: $C_{NL}(\sigma) > C_L(\sigma)$. Upon whether a leniency program raises or lowers the frequency of cartels then comes down to its impact on non-leniency enforcement. If a leniency program strengthens non-leniency enforcement - $\sigma_L^* > \sigma_{NL}^*$ - then clearly a leniency program lowers the cartel rate: $C_L(\sigma_L^*) < C_{NL}(\sigma_{NL}^*)$. If a leniency program weakens non-leniency enforcement - $\sigma_L^* < \sigma_{NL}^*$ - then the ultimate impact on the cartel rate is ambiguous. However, if a leniency program has a small effect on the cartel rate holding non-leniency enforcement fixed - $C_{NL}(\sigma) \simeq C_L(\sigma)$ - but significantly undermines non-leniency enforcement - $\sigma_L^* << \sigma_{NL}^*$ - then a leniency program will result in more cartels.

Recall that the probability a cartel pays penalties through non-leniency enforcement is given by

$$\sigma_{NL}^* = qrp\left(qr\int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma_{NL}^*, \eta\right) g(\eta)\, d\eta\right)$$

when there is no leniency program and

$$\sigma_L^* = qrp\left(\lambda\int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right)\right) C_L\left(\sigma_L^*, \eta\right) g(\eta)\, d\eta + qr\int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right) C_L\left(\sigma_L^*, \eta\right) g(\eta)\, d\eta\right)$$

when there is a leniency program. Without a leniency program, the caseload for a CA is given by $qr\int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma_{NL}^*, \eta\right) g(\eta)\, d\eta$ where $\int_{\underline{\eta}}^{\overline{\eta}} C\left(\sigma_{NL}^*, \eta\right) g(\eta)\, d\eta$ is the mass of cartels, q is the fraction of cartels discovered, and r is the fraction of discovered cartels that are prosecuted. With a leniency program, cases come from dying cartels,

$$\int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right)\right) C_L\left(\sigma_L^*, \eta\right) g(\eta)\, d\eta,$$

and from cartels discovered and prosecuted (and which did not collapse and thus apply for leniency),

$$qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right) C_L\left(\sigma_L^*, \eta\right) g\left(\eta\right) d\eta.$$

As leniency cases only use up λ as many resources as one non-leniency case, the total caseload is

$$\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right)\right) C_L\left(\sigma_L^*, \eta\right) g\left(\eta\right) d\eta + qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_L^*, \eta\right)\right) C_L\left(\sigma_L^*, \eta\right) g\left(\eta\right) d\eta.$$

To begin, let us consider what would happen if the source of cases was unchanged with a leniency program in that it is a fraction qr of all cartels. Given that the cartel rate is lower with a leniency program holding σ fixed, if, given non-leniency enforcement is at its level when there is no leniency program, then caseload is smaller with a leniency program because there are fewer cartels:

$$qr \int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta > qr \int_{\underline{\eta}}^{\overline{\eta}} C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta.$$

A smaller caseload would then imply a higher value of σ which would feedback to result in a lower cartel rate which would reduce caseload more, raise σ more, and so forth. In that case, a lower cartel rate (holding σ fixed) is reinforced when σ is endogenized. Of course, the impact on caseload from a leniency program is different from just described in two ways. First, holding the cartel rate fixed, there are more cases because all dying cartels becomes cases by applying for leniency while, without a leniency program, only a fraction qr become cases by being discovered and prosecuted. These additional cases will, ceteris paribus, lower σ and thus feedback to raise the cartel rate. Second, if $\lambda < 1$ then leniency cases take up fewer resources. Without a leniency program, some dying cartels were discovered and prosecuted. With a leniency program, those cartels become leniency applicants and there is a savings of $1-\lambda$ for each case in terms of caseload. This effect reduces caseload which raises σ and thus feedbacks to lower the cartel rate. Summing up, if either a leniency program does not produce many applications or a leniency application takes up sufficiently few resources then the smaller caseload will enhance non-leniency enforcement and result in more intense non-leniency enforcement, $\sigma_L^* > \sigma_{NL}^*$, and a lower cartel rate with a leniency program, $C_L\left(\sigma_L^*\right) < C_{NL}\left(\sigma_{NL}^*\right)$. If, however, there are many leniency applications and those applications do require substantial resources then non-leniency enforcement could be weakened, $\sigma_L^* < \sigma_{NL}^*$, because leniency cases crowd out non-leniency cases and, therefore, a cartel assigns a lower probability to being convicted in the event of discovery and prosecution. In that case, the impact on the cartel rate of introducing a leniency program is unclear.

With the preceding discussion, we can now better interpret our findings. Theorems 7 and 8 provide sufficient conditions for a leniency program to reduce the cartel rate and, in both cases, it is because a leniency program does not significantly add to caseload. With Theorem 7, the assumption of a uniform distribution on market conditions has the implication that cartels never internally collapse; a cartel is either stable for all market conditions or no market conditions (in which case it does not form). Given that, in equilibrium, firms apply for leniency only when a cartel collapses, there are then no leniency applications; hence, a leniency program does not crowd out non-leniency cases and, as a result, does not weaken non-leniency enforcement. In fact, it strengthens non-leniency enforcement. Though leniency is not applied for in equilibrium, the possibility of doing so makes deviation more attractive which serves to tighten the ICC and thereby produce fewer cartels. Given that the cartel rate is lower, there are fewer non-leniency cases which raises the conviction rate and thus enhances non-leniency enforcement. With Theorem 8, λ is low so that leniency cases do not require too much in terms of resources. As a result, even with firms applying for leniency, the caseload is not higher with a leniency program so a lower cartel rate emerges, both from the cartel-destabilizing effect of a leniency program - fewer industries are able to cartelize and those that do have shorter duration - and the strengthened non-leniency enforcement - the fewer number of non-leniency cases due to a lower cartel rate dominates the additional cases from the leniency program. Similarly, Theorem 9 reduces the cartel rate because it does not weaken non-leniency enforcement though, in that situation, it is because non-leniency enforcement is initially weak.

Theorem 10 shows that when penalties are sufficiently weak and leniency cases do not take fewer resources than non-leniency cases then the cartel rate is higher with a leniency program. Given that penalties are positive (though small), it is still the case that the cartel rate is lower with a leniency program holding fixed non-leniency enforcement: $C_L(\sigma^*_{NL}) < C_{NL}(\sigma^*_{NL})$. However, what happens is that non-leniency enforcement is sufficiently weakened by a leniency program - $\sigma^*_L << \sigma^*_{NL}$ - that it overwhelms the lower cartel rate function so that $C_L(\sigma^*_L) > C_{NL}(\sigma^*_{NL})$. All this is driven by the fact that leniency cases crowd out non-leniency cases. As leniency cases come from cartels that have stopped operating anyway, leniency cases do not promote desistance; in contrast, successful non-leniency cases shut down active cartels and thus promote desistance. Prosecuting all dying cartels for sure - which is what occurs with a leniency program - rather than only with probability qr (as without a leniency program) can promote deterrence by increasing expected penalties. However, if penalties are small then this effect is overwhelmed by the shift of CA resources to prosecuting leniency cases and the subsequent weakening of non-leniency enforcement. This result highlights how penalties are a critical complement to a leniency program. If penalties are weak then a leniency program may not just be ineffective but rather counter-productive as it can raise the cartel rate because of its deleterious effect on non-leniency enforcement. In the next section, numerical analysis will allow us to tell a richer story of the potential counter-productive impact on non-leniency enforcement of a leniency program.

6 Impact of a Leniency Program: Numerical Results (To Be Done)

7 Concluding Remarks (To Be Done)

8 Appendix

Proof of Lemma 1. ψ can be presented as:

$$\begin{aligned}\psi(Y) &= \int_{\underline{\pi}}^{\phi(Y)} \left[(1-\delta)\pi + \delta Y - \delta\sigma(Y-W) - (1-\delta)\sigma\gamma(Y-\alpha\mu)\right] h(\pi)\, d\pi \\ &\quad + \int_{\phi(Y)}^{\overline{\pi}} \left[(1-\delta)\alpha\pi + \delta W - (1-\delta)\zeta(\theta,\sigma)\gamma(Y-\alpha\mu)\right] h(\pi)\, d\pi\end{aligned} \tag{14}$$

where

$$\phi(Y,\sigma,\eta) = \frac{\delta(1-\sigma)(1-\kappa)(Y-\alpha\mu) - [1-\delta(1-\kappa)]\left[\sigma - \min\{\sigma,\theta\}\right]\gamma(Y-\alpha\mu)}{(\eta-1)[1-\delta(1-\kappa)]} \tag{15}$$

and

$$\zeta(\theta,\sigma) = \begin{cases} \sigma & \text{if } \sigma \le \theta \\ \omega & \text{if } \sigma > \theta \end{cases} \tag{16}$$

Evaluate (14) at $Y = \mu$ and set $W = \mu$ which gives an upper bound on $\psi(\mu)$ since $W \le Y$ and ψ is increasing in W:

$$\begin{aligned}\psi(\mu) &\le \int_{\underline{\pi}}^{\phi(\mu)} \left[(1-\delta)\pi + \delta\mu - (1-\delta)\sigma\gamma\mu(1-\alpha)\right] h(\pi)\, d\pi \\ &\quad + \int_{\phi(\mu)}^{\overline{\pi}} \left[(1-\delta)\alpha\pi + \delta\mu - (1-\delta)\zeta(\theta,\sigma)\gamma\mu(1-\alpha)\right] h(\pi)\, d\pi\end{aligned}$$

$$\begin{aligned}\psi(\mu) &\le \int_{\underline{\pi}}^{\phi(\mu)} \left[(1-\delta)\pi + \delta\mu\right] h(\pi)\, d\pi - \int_{\underline{\pi}}^{\phi(\mu)} (1-\delta)\sigma\gamma\mu(1-\alpha)\, h(\pi)\, d\pi \\ &\quad + \int_{\phi(\mu)}^{\overline{\pi}} \left[(1-\delta)\alpha\pi + \delta\mu - (1-\delta)\zeta(\theta,\sigma)\gamma\mu(1-\alpha)\right] h(\pi)\, d\pi\end{aligned}$$

$$\begin{aligned}\psi(\mu) &\le \delta\mu + \int_{\underline{\pi}}^{\phi(\mu)} (1-\delta)\pi h(\pi)\, d\pi + \int_{\phi(\mu)}^{\overline{\pi}} (1-\delta)\alpha\pi h(\pi)\, d\pi \\ &\quad - \int_{\underline{\pi}}^{\phi(\mu)} (1-\delta)\sigma\gamma\mu(1-\alpha)\, h(\pi)\, d\pi - \int_{\phi(\mu)}^{\overline{\pi}} (1-\delta)\zeta(\theta,\sigma)\gamma\mu(1-\alpha)\, h(\pi)\, d\pi\end{aligned}$$

$$\begin{aligned}\psi(\mu) &\leq \delta\mu + \int_{\underline{\pi}}^{\overline{\pi}} (1-\delta)\,\pi h(\pi)\,d\pi - \int_{\phi(\mu)}^{\overline{\pi}} (1-\delta)(1-\alpha)\,\pi h(\pi)\,d\pi \\ &\quad - \int_{\underline{\pi}}^{\phi(\mu)} (1-\delta)\,\sigma\gamma\mu(1-\alpha)\,h(\pi)\,d\pi - \int_{\phi(\mu)}^{\overline{\pi}} (1-\delta)\,\zeta(\theta,\sigma)\,\gamma\mu(1-\alpha)\,h(\pi)\,d\pi\end{aligned}$$

$$\begin{aligned}\psi(\mu) &\leq \mu - \int_{\phi(\mu)}^{\overline{\pi}} (1-\delta)(1-\alpha)\,\pi h(\pi)\,d\pi - \int_{\underline{\pi}}^{\phi(\mu)} (1-\delta)\,\sigma\gamma\mu(1-\alpha)\,h(\pi)\,d\pi \\ &\quad - \int_{\phi(\mu)}^{\overline{\pi}} (1-\delta)\,\zeta(\theta,\sigma)\,\gamma\mu(1-\alpha)\,h(\pi)\,d\pi.\end{aligned}$$

Therefore, $\psi(\mu) \leq \mu$. Given that $\psi(\alpha\mu) = \alpha\mu$, it follows that ψ maps $[\alpha\mu, \mu]$ into itself as long as $\psi'(Y) \geq 0$. We'll now show that property holds.

To prove $\psi'(Y) \geq 0$ when $\gamma \simeq 0$, begin with when $\phi(Y) < \underline{\pi}$ so that (14) takes the form:

$$\psi(Y) = \int_{\underline{\pi}}^{\overline{\pi}} \left[(1-\delta)\,\alpha\pi + \delta W - (1-\delta)\,\zeta(\theta,\sigma)\,\gamma(Y-\alpha\mu)\right] h(\pi)\,d\pi.$$

Thus,

$$\begin{aligned}\psi'(Y) &= \int_{\underline{\pi}}^{\overline{\pi}} \left[\delta\frac{\partial W}{\partial Y} - (1-\delta)\,\zeta(\theta,\sigma)\,\gamma\right] h(\pi)\,d\pi \\ &= \frac{\delta\kappa}{1-\delta(1-\kappa)} - (1-\delta)\,\zeta(\theta,\sigma)\,\gamma\end{aligned}$$

and $\psi'(Y) \in (0,1)$ when $\gamma \simeq 0$. If $\overline{\pi} < \phi(Y)$ then (14) takes the form:

$$\psi(Y) = \int_{\underline{\pi}}^{\overline{\pi}} \left[(1-\delta)\,\pi + \delta Y - \delta\sigma(Y-W)\right] h(\pi)\,d\pi - (1-\delta)\,\sigma\gamma(Y-\alpha\mu),$$

and

$$\psi'(Y) = \delta\left[1-\sigma\left(1-\frac{\partial W}{\partial Y}\right) - (1-\delta)\,\sigma\gamma\right] = \delta\left[(1-\sigma)\left(1-\frac{\partial W}{\partial Y}\right) + \frac{\partial W}{\partial Y}\right] - (1-\delta)\,\sigma\gamma.$$

Given that $\partial W/\partial Y = \kappa/(1-\delta(1-\kappa)) \in (0,1)$, if $\gamma \simeq 0$ then $\psi'(Y) \in (0,1)$. Finally, if $\phi(Y) \in (\underline{\pi}, \overline{\pi})$ then differentiating (14) yields:

$$\begin{aligned}\psi'(Y) &= \left[(1-\delta)\,\phi(Y) + \delta Y - \delta\sigma(Y-W) - (1-\delta)\,\sigma\gamma(Y-\alpha\mu)\right] h(\phi(Y))\,\phi'(Y) \\ &\quad + \int_{\underline{\pi}}^{\phi(Y)} \left\{\delta\left[1-\sigma\left(1-\frac{\partial W}{\partial Y}\right)\right] - (1-\delta)\,\sigma\gamma\right\} h(\pi)\,d\pi \\ &\quad - \left[(1-\delta)\,\alpha\phi(Y) + \delta W - (1-\delta)\,\zeta(\theta,\sigma)\,\gamma(Y-\alpha\mu)\right] h(\phi(Y))\,\phi'(Y) \\ &\quad + \int_{\phi(Y)}^{\overline{\pi}} \left[\delta\frac{\partial W}{\partial Y} - (1-\delta)\,\zeta(\theta,\sigma)\,\gamma\right] h(\pi)\,d\pi\end{aligned}$$

$$\begin{aligned}\psi'(Y) &= [(1-\delta)(1-\alpha)\phi(Y)+\delta(1-\sigma)(Y-W)]h(\phi(Y))\phi'(Y) \\ &+\int_{\underline{\pi}}^{\phi(Y)}\delta(1-\sigma)\left(1-\frac{\partial W}{\partial Y}\right)h(\pi)\,d\pi+\delta\frac{\partial W}{\partial Y} \\ &+\gamma(1-\delta)\left[(\zeta(\theta,\sigma)-\sigma)(Y-\alpha\mu)h(\phi(Y))\phi'(Y)-\sigma H(\phi(Y))+\zeta(\theta,\sigma)(1-H(\phi(Y)))\right]\end{aligned} \tag{17}$$

In evaluating the sign of $\psi'(Y)$, note that if $\gamma \simeq 0$ then

$$\phi'(Y)=\frac{\delta(1-\sigma)(1-\kappa)}{(1-\delta(1-\kappa))(\eta-1)}-\frac{[1-\delta(1-\kappa)][\sigma-\min\{\sigma,\theta\}]\gamma}{\eta-1}>0.$$

(17) sums up four terms. The first term is positive. Next note that $\partial W/\partial Y \in (0,1)$ implies the second and third terms are positive. If $\gamma \simeq 0$ then the fourth term is small relative to the first three terms which implies $\psi'(Y)>0$. ■

Proof of Lemma 2. First, let us consider the impact on $Y^*(\sigma,\eta)$ of changing σ. Initially, let us suppose that $\phi^*(\sigma,\eta)\in(\underline{\pi},\overline{\pi})$. Referring to (14)-(16) and given that $\omega>\theta$, there is a discontinuous decrease in $\psi(Y)$ at $\sigma=\theta$ when σ is increased as the penalty jumps from $\theta\gamma(Y-\alpha\mu)$ to $\omega\gamma(Y-\alpha\mu)$. Thus, at $\sigma=\theta$, $\psi(Y)$ is decreasing in σ. Next consider the response of $\psi(Y)$ to σ when $\sigma\neq\theta$ and thus is differentiable:

$$\begin{aligned}\frac{\partial\psi(Y)}{\partial\sigma} &= [(1-\delta)\phi+\delta Y-\delta\sigma(Y-W)-(1-\delta)\sigma\gamma(Y-\alpha\mu)]h(\phi)\frac{\partial\phi}{\partial\sigma} \\ &-\int_{\underline{\pi}}^{\phi(Y)}[\delta(Y-W)+(1-\delta)\gamma(Y-\alpha\mu)]h(\pi)\,d\pi \\ &-[(1-\delta)\alpha\phi+\delta W-(1-\delta)\zeta(\theta,\sigma)\gamma(Y-\alpha\mu)]h(\phi)\frac{\partial\phi}{\partial\sigma} \\ &-\int_{\phi(Y)}^{\overline{\pi}}(1-\delta)\frac{\partial\zeta(\theta,\sigma)}{\partial\sigma}\gamma(Y-\alpha\mu)h(\pi)\,d\pi\end{aligned}$$

$$\begin{aligned}\frac{\partial\psi(Y)}{\partial\sigma} &= [(1-\delta)(1-\alpha)\phi+\delta(1-\sigma)(Y-W)]h(\phi)\frac{\partial\phi}{\partial\sigma} \\ &-\int_{\underline{\pi}}^{\phi(Y)}[\delta(Y-W)+(1-\delta)\gamma(Y-\alpha\mu)]h(\pi)\,d\pi \\ &-\int_{\phi(Y)}^{\overline{\pi}}(1-\delta)\frac{\partial\zeta(\theta,\sigma)}{\partial\sigma}\gamma(Y-\alpha\mu)h(\pi)\,d\pi \\ &+(1-\delta)(\zeta(\theta,\sigma)-\sigma)\gamma(Y-\alpha\mu)h(\phi)\frac{\partial\phi}{\partial\sigma}\end{aligned} \tag{18}$$

where

$$\frac{\partial\zeta(\theta,\sigma)}{\partial\sigma}=\begin{cases}1 & \text{if } \sigma<\theta\\ 0 & \text{if } \sigma>\theta\end{cases}$$

In signing these terms, recall that we are focusing on the case of $\phi^*(\sigma,\eta)\in(\underline{\pi},\overline{\pi})$ which implies $Y^*(\sigma,\eta)>\alpha\mu$. Given that

$$\frac{\partial\phi}{\partial\sigma}=-\left(\frac{\delta(1-\kappa)(Y-\alpha\mu)}{(1-\delta(1-\kappa))(\eta-1)}\right)-\frac{\mathbf{1}_{\sigma>\theta}\gamma(Y-\alpha\mu)}{\eta-1}<0,$$

where $\mathbf{1}_{\sigma>\theta} = 1$ if $\sigma > \theta$ and 0 otherwise, then the first term in (18) is negative. The second term is negative and the third term is non-positive. If $\gamma \simeq 0$ then the fourth term is small relative to the first two terms from which we can conclude $\partial\psi(Y)/\partial\sigma < 0$. Now consider increasing σ from σ' to σ''. Given that $\psi(Y)$ is decreasing in σ then $\psi(Y)$ shifts down. Since $\psi(Y,\sigma') \leq Y$ as $Y \geq Y^*(\sigma',\eta)$ (recalling that Y^* is the maximal fixed point) then $\psi(Y,\sigma'') < Y$ for all $Y \geq Y^*(\sigma',\eta)$ which implies $Y^*(\sigma'',\eta) < Y^*(\sigma',\eta)$. Thus, if $\phi^*(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$, which implies $Y^*(\sigma,\eta) > \alpha\mu$, then $Y^*(\sigma,\eta)$ is decreasing in σ.

Next suppose $\phi^*(\sigma,\eta) = \overline{\pi}$ in which case

$$\psi(Y) = \int_{\underline{\pi}}^{\overline{\pi}} \left[(1-\delta)\pi + \delta Y - \delta\sigma(Y-W) - (1-\delta)\sigma\gamma(Y-\alpha\mu)\right] h(\pi)\,d\pi$$

and

$$\frac{\partial\psi(Y)}{\partial\sigma} = -\int_{\underline{\pi}}^{\overline{\pi}} \left[\delta(Y-W) + (1-\delta)\gamma(Y-\alpha\mu)\right] h(\pi)\,d\pi < 0.$$

By the previous argument, it follows that $Y^*(\sigma,\eta)$ is decreasing in σ. Note that if $\phi^*(\sigma,\eta) = \overline{\pi}$ then $Y^*(\sigma,\eta) > \alpha\mu$.

Finally, suppose $\phi^*(\sigma,\eta) = \underline{\pi}$. Given that $Y^*(\sigma,\eta) = \alpha\mu$ then $Y^*(\sigma,\eta)$ is independent of σ. Summing up, it has been shown that $Y^*(\sigma,\eta)$ is non-increasing in σ and, when $Y^*(\sigma,\eta) > \alpha\mu$, $Y^*(\sigma,\eta)$ is decreasing in σ.

Now consider the impact on $Y^*(\sigma,\eta)$ of changing η. Note that η only operates through ϕ since, in equilibrium, a firm never cheats. Hence, if $\phi^*(\sigma,\eta) \in \{\underline{\pi},\overline{\pi}\}$ then $Y^*(\sigma,\eta)$ is independent of η. Let us then suppose $\phi^*(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$ in which case

$$\begin{aligned}\psi(Y) &= \int_{\underline{\pi}}^{\phi(Y)} \left[(1-\delta)\pi + \delta Y - \delta\sigma(Y-W) - (1-\delta)\sigma\gamma(Y-\alpha\mu)\right] h(\pi)\,d\pi \\ &\quad + \int_{\phi(Y)}^{\overline{\pi}} \left[(1-\delta)\alpha\pi + \delta W - (1-\delta)\zeta(\theta,\sigma)\gamma(Y-\alpha\mu)\right] h(\pi)\,d\pi\end{aligned}$$

at $Y = Y^*(\sigma,\eta)$. We then have

$$\begin{aligned}\frac{\partial\psi}{\partial\eta} &= \{[(1-\delta)\phi + \delta Y - \delta\sigma(Y-W) - (1-\delta)\sigma\gamma(Y-\alpha\mu)] \\ &\qquad -(1-\delta)\alpha\phi - \delta W + (1-\delta)\zeta(\theta,\sigma)\gamma(Y-\alpha\mu)\} h(\phi)\left(\frac{\partial\phi}{\partial\eta}\right) \\ &= \{[(1-\delta)(1-\alpha)\phi + \delta(1-\sigma)(Y-W) + (1-\delta)(\zeta(\theta,\sigma)-\sigma)\gamma(Y-\alpha\mu)]\} h(\phi)\left(\frac{\partial\phi}{\partial\eta}\right)\end{aligned}$$

If $\gamma \simeq 0$ then the expression in $\{\}$ is positive in which case

$$sign\left\{\frac{\partial\psi}{\partial\eta}\right\} = sign\left\{\frac{\partial\phi}{\partial\eta}\right\}.$$

Given that

$$\frac{\partial\phi}{\partial\eta} = -\left(\frac{\delta(1-\sigma)(1-\kappa)(Y-\alpha\mu) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma,\theta\}]\gamma(Y-\alpha\mu)}{(\eta-1)^2[1-\delta(1-\kappa)]}\right) = -\frac{\phi}{\eta-1} < 0$$

then $\partial \psi / \partial \eta < 0$. Hence, raising η lowers $\psi(Y, \sigma, \eta)$ and thus lowers $Y^*(\sigma, \eta)$. We have then shown that $Y^*(\sigma, \eta)$ is non-increasing in η and, when $\phi^*(\sigma, \eta) \in (\underline{\pi}, \overline{\pi})$, $Y^*(\sigma, \eta)$ is decreasing in η.

Turning to comparative statics for $\phi^*(\sigma, \eta)$, recall that

$$\phi^*(\sigma, \eta) \equiv \max\left\{\min\left\{\phi\left(Y^*(\sigma, \eta), \sigma, \eta\right), \overline{\pi}\right\}, \underline{\pi}\right\}$$

where the expression for $\phi(Y, \sigma, \eta)$ is in (15). If $\phi^*(\sigma, \eta) \in \{\underline{\pi}, \overline{\pi}\}$ then it is independent of σ and η. Thus, from hereon assume $\phi^*(\sigma, \eta) \in (\underline{\pi}, \overline{\pi})$. In that case,

$$\frac{\partial \phi\left(Y^*(\sigma, \eta), \sigma, \eta\right)}{\partial Y} = \frac{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma, \theta\}]\gamma}{(\eta-1)[1-\delta(1-\kappa)]}$$

and

$$\frac{\partial \phi\left(Y^*(\sigma, \eta), \sigma, \eta\right)}{\partial \sigma} = \begin{cases} -\left(\frac{\delta(1-\kappa)(Y^*(\sigma,\eta)-\alpha\mu)+[1-\delta(1-\kappa)]\gamma(Y^*(\sigma,\eta)-\alpha\mu)}{(\eta-1)[1-\delta(1-\kappa)]}\right) & \text{if } \sigma > \theta \\ -\frac{\delta(1-\kappa)(Y^*(\sigma,\eta)-\alpha\mu)}{(\eta-1)[1-\delta(1-\kappa)]} & \text{if } \sigma < \theta \end{cases}.$$

$\phi(Y, \sigma, \eta)$ is then decreasing in σ and, when $\gamma \simeq 0$, increasing in Y. Given that it has already been shown that $Y^*(\sigma, \eta)$ is decreasing in σ, it follows that $\phi^*(\sigma, \eta)$ is decreasing in σ.

Finally, consider the comparative static of $\phi^*(\sigma, \eta)$ with respect to η. If $\phi^*(\sigma, \eta) \in (\underline{\pi}, \overline{\pi})$ then re-arranging (15) yields

$$\phi(Y, \sigma, \eta) = \left(\frac{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma, \theta\}]\gamma}{1-\delta(1-\kappa)}\right)\left(\frac{Y-\alpha\mu}{\eta-1}\right).$$

Consider $\eta'' > \eta'$,

$$\begin{aligned} \phi^*\left(\sigma, \eta''\right) - \phi^*\left(\sigma, \eta'\right) &= \left(\frac{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma, \theta\}]\gamma}{1-\delta(1-\kappa)}\right) \times \\ &\quad \left[\left(\frac{Y^*(\sigma, \eta'') - \alpha\mu}{\eta''-1}\right) - \left(\frac{Y^*(\sigma, \eta') - \alpha\mu}{\eta''-1}\right)\right]. \end{aligned}$$

If $\gamma \simeq 0$ then

$$\frac{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma, \theta\}]\gamma}{1-\delta(1-\kappa)} > 0$$

and, therefore,

$$sign\left\{\phi^*\left(\sigma, \eta''\right) - \phi^*\left(\sigma, \eta'\right)\right\} = sign\left\{\left(\frac{Y^*(\sigma, \eta'') - \alpha\mu}{\eta''-1}\right) - \left(\frac{Y^*(\sigma, \eta') - \alpha\mu}{\eta'-1}\right)\right\}.$$

This expression is negative because $\eta'' - 1 > \eta' - 1 > 0$ and, given that $Y^*(\sigma, \eta)$ is decreasing in η, $Y^*(\sigma, \eta') - \alpha\mu > Y^*(\sigma, \eta'') - \alpha\mu > 0$. This proves $\phi^*(\sigma, \eta)$ is decreasing in η. ■

Proof of Lemma 3. [The method of proof is the same as that used in proving Theorem 4 in Harrington and Chang (2009).] Recall that $\widehat{\eta}(\sigma)$ satisfies the property,

$$Y^*(\sigma,\eta)\begin{cases} >\alpha\mu & \text{if } \eta\in\left[\underline{\eta},\widehat{\eta}(\sigma)\right] \\ =\alpha\mu & \text{if } \eta>\widehat{\eta}(\sigma) \end{cases}$$

Suppose $\widehat{\eta}(\sigma)\in\left(\underline{\eta},\overline{\eta}\right)$. By the continuity of $\psi(Y,\sigma,\eta)$ with respect to Y and η and that $\psi(Y,\sigma,\eta)$ is non-increasing in η (see the proof of Lemma 2), it follows that

$$\psi(Y,\widehat{\eta}(\sigma))\leq Y,\ \text{for all } Y,\ \text{and } \psi(Y^*(\sigma,\widehat{\eta}(\sigma)),\widehat{\eta}(\sigma))=Y^*(\sigma,\widehat{\eta}(\sigma)).$$

With this property, let us now argue that $\widehat{\eta}$ is decreasing in σ. Consider raising σ from σ' to σ''. Since $\psi(Y,\eta)$ is decreasing in σ for $Y>\alpha\mu$ (see the proof of Lemma 2) then

$$\psi\left(Y,\widehat{\eta}\left(\sigma'\right),\sigma''\right)<Y,\ \text{for all } Y>\alpha\mu.$$

Given that $\psi(Y,\eta)$ is decreasing in η for $Y>\alpha\mu$ then $\widehat{\eta}(\sigma'')<\widehat{\eta}(\sigma')$. ■

Proof of Lemma 4. [The method of proof is the same as that used in proving in Harrington and Chang (2009).] Recall that the cartel rate for a type-η industry is

$$C(\sigma,\eta)=\frac{\kappa(1-\sigma)H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi^*(\sigma,\eta))}.$$

Suppose σ is raised from σ' to σ''. The change in the type-η industry cartel rate is

$$\frac{\kappa(1-\sigma'')H(\phi^*(\sigma'',\eta))}{1-(1-\kappa)(1-\sigma'')H(\phi^*(\sigma'',\eta))}-\frac{\kappa(1-\sigma')H(\phi^*(\sigma',\eta))}{1-(1-\kappa)(1-\sigma')H(\phi^*(\sigma',\eta))}.$$

The sign of that expression is the same as

$$\begin{aligned} & sign\left\{\kappa\left(1-\sigma''\right)H\left(\phi^*\left(\sigma'',\eta\right)\right)\left[1-(1-\kappa)\left(1-\sigma'\right)H\left(\phi^*\left(\sigma',\eta\right)\right)\right]-\right. \\ & \left.\kappa\left(1-\sigma'\right)H\left(\phi^*\left(\sigma',\eta\right)\right)\left[1-(1-\kappa)\left(1-\sigma''\right)H\left(\phi^*\left(\sigma'',\eta\right)\right)\right]\right\} \\ = \ & sign\left\{\left(1-\sigma''\right)H\left(\phi^*\left(\sigma'',\eta\right)\right)-\left(1-\sigma'\right)H\left(\phi^*\left(\sigma',\eta\right)\right)\right\}<0. \end{aligned}$$

This expression is negative because $\sigma''>\sigma'$ implies $1-\sigma''<1-\sigma'$, and $H(\phi^*(\sigma'',\eta))<H(\phi^*(\sigma',\eta))$ because $\phi^*(\sigma'',\eta)<\phi^*(\sigma',\eta)$ (by Lemma 2).

Next consider the aggregate cartel rate,

$$\begin{aligned} C(\sigma) &= \int_{\underline{\eta}}^{\overline{\eta}}\left[\frac{\kappa(1-\sigma)H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi^*(\sigma,\eta))}\right]g(\eta)\,d\eta \\ &= \int_{\underline{\eta}}^{\widehat{\eta}(\sigma)}\left[\frac{\kappa(1-\sigma)H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi^*(\sigma,\eta))}\right]g(\eta)\,d\eta \end{aligned}$$

It was just shown that the integrand is decreasing in σ and, since $\widehat{\eta}(\sigma)$ is decreasing in σ by Lemma 3, this expression is decreasing in σ. ■

Proof of Theorem 5. When $\theta = 1$ then

$$\Psi(\sigma) = qrp\left(qr\int_{\underline{\eta}}^{\overline{\eta}} C(\sigma,\eta)\, g(\eta)\, d\eta\right), \tag{19}$$

and when $\theta = 0$ then

$$\begin{aligned}\Psi(\sigma) \quad &= \quad qrp\left(\lambda\int_{\underline{\eta}}^{\overline{\eta}} (1 - H(\phi^*(\sigma,\eta)))\, C(\sigma,\eta)\, g(\eta)\, d\eta \right. \qquad (20)\\ &\qquad \left. +qr\int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma,\eta))\, C(\sigma,\eta)\, g(\eta)\, d\eta\right).\end{aligned}$$

To show that a fixed point exists for (19) and for (20), the proof strategy has two steps: 1) show that, for any value of σ, the integrand in these equations is continuous in σ except for a countable set of values of η; and 2) show that it follows from step 1 that Ψ is continuous. The proof will focus exclusively on proving that (20) has a fixed point as the method of proof is immediately applicable to the case of (19).[25]

Considering the integrand in (20), a discontinuity in

$$H(\phi^*(\sigma,\eta))\, C(\sigma,\eta)\, g(\eta) = H(\phi^*(\sigma,\eta))\left(\frac{\kappa(1-\sigma)\, H(\phi^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)\, H(\phi^*(\sigma,\eta))}\right) g(\eta)$$

with respect to σ (or η) comes from $\phi^*(\sigma,\eta)$ being discontinuous, which comes from $Y^*(\sigma,\eta)$ being discontinuous. Let $\Delta(\sigma') \subseteq [\underline{\eta},\overline{\eta}]$ be the set of η for which $Y^*(\sigma,\eta)$ is discontinuous at $\sigma = \sigma'$. We will show that $\Delta(\sigma)$ is countable.

Suppose $Y^*(\sigma,\eta)$ is discontinuous in σ at $(\sigma,\eta) = (\sigma',\eta')$. Given $\psi(Y,\sigma,\eta)$ is continuous and $Y^*(\sigma,\eta)$ is the maximal fixed point to $\psi(Y,\sigma,\eta)$ then

$$\psi(Y,\sigma',\eta') < Y, \forall Y \in \left(Y^*(\sigma',\eta'), \mu\right]. \tag{21}$$

If, in addition, $\exists \xi > 0$ such that

$$\psi(Y,\sigma',\eta') > Y, \forall Y \in \left[Y^*(\sigma',\eta') - \xi, Y^*(\sigma',\eta')\right)$$

then, by the continuity of $\psi(Y,\sigma,\eta)$ in σ, $Y^*(\sigma,\eta)$ is continuous at $(\sigma,\eta) = (\sigma',\eta')$, contrary to our supposition. Hence, it must be the case that $\exists \xi > 0$ such that

$$\psi(Y,\sigma',\eta') \leq Y, \forall Y \in \left[Y^*(\sigma',\eta') - \xi, Y^*(\sigma',\eta')\right]. \tag{22}$$

Figures 1a-1c cover the possible cases in which Y^* is discontinuous.

Insert Figures 1a-1c here

[25] When $\theta = 1$, existence of a fixed point can also be established by showing that $\Psi(\sigma)$ is non-decreasing in σ and appealing to Tarski's Fixed Point Theorem. However, when $\theta < 1$, it is generally not true that $\Psi(\sigma)$ is non-decreasing in σ $\forall\sigma$.

Given that $\psi(Y, \sigma, \eta)$ is continuous and decreasing in η (see the proof of Lemma 2) then (21) and (22) imply

$$\psi\left(Y, \sigma', \eta'\right) < Y, \forall Y \in \left[Y^*\left(\sigma', \eta'\right) - \xi, \mu\right], \ \forall \eta > \eta'. \tag{23}$$

It follows from (23) that, $\forall \eta > \eta'$, all fixed points to ψ are bounded above by $Y^*(\sigma', \eta') - \xi$:

$$Y^*\left(\sigma', \eta\right) < Y^*\left(\sigma', \eta'\right) - \xi, \ \forall \eta > \eta'.$$

Next define:

$$\varepsilon\left(\sigma', \eta'\right) \equiv Y^*\left(\sigma', \eta'\right) - \lim_{\eta \downarrow \eta'} Y^*\left(\sigma', \eta\right)$$

where $\varepsilon(\sigma', \eta')$ measures the size of the discontinuity in $Y^*(\sigma', \eta)$ with respect to η at $\eta = \eta'$; see Figure 2.

Insert Figure 2 here

For each $\eta \in \Delta(\sigma')$, there has then been associated an interval of length $\varepsilon(\sigma', \eta)$. Note that these intervals have a null intersection because $Y^*(\sigma, \eta)$ is non-increasing in η. Hence,

$$\sum_{\eta \in \Delta(\sigma')} \varepsilon\left(\sigma', \eta\right) \leq (1 - \alpha)\mu.$$

Given that a sum can only be finite if the number of elements which are positive is countable, it follows that $\Delta(\sigma')$ is countable. Hence, the set of values for η for which $Y^*(\sigma', \eta)$ is discontinuous in σ at $\sigma = \sigma'$ is countable. This completes the first step.

By Jeffrey (1925), given that $H(\phi^*(\sigma, \eta)) C(\sigma, \eta) g(\eta)$ and $(1 - H(\phi^*(\sigma, \eta))) C(\sigma, \eta) g(\eta)$ are bounded in (σ, η) on $[0, 1] \times \left[\underline{\eta}, \overline{\eta}\right]$ and are continuous at $\sigma = \sigma'$ for all $\eta \in \left[\underline{\eta}, \overline{\eta}\right]$ except for a countable set then

$$\int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma, \eta)) C(\sigma, \eta) g(\eta) d\eta$$

and

$$\int_{\underline{\eta}}^{\overline{\eta}} (1 - H(\phi^*(\sigma, \eta))) C(\sigma, \eta) g(\eta) d\eta$$

are continuous at $\sigma = \sigma'$. Given that p is a continuous function, it follows that

$$p\left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} (1 - H(\phi^*(\sigma, \eta))) C(\sigma, \eta) g(\eta) d\eta + qr \int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma, \eta)) C(\sigma, \eta) g(\eta) d\eta\right)$$

is continuous in σ. Hence, Ψ in (20) is continuous in σ and maps $[0, 1]$ into itself; therefore, a fixed point exists. The same method of proof can be used to show that a fixed point to (19) exists. ■

Proof of Theorem 6. The proof has three steps. First, holding Y fixed, the threshold for stable collusion is shown to be lower with a leniency program: $\phi_{NL}(Y, \eta) > \phi_\theta(Y, \eta)$. When $\sigma > \theta$, which holds by supposition, the deviator has lower penalties by applying for leniency and this tightens the ICC and thus raises the threshold. Second, given $\phi_{NL}(Y,\eta) > \phi_\theta(Y,\eta)$ and the supposition that $\omega > \sigma$, it is shown that $\psi_{NL}(Y,\sigma,\eta) > \psi_\theta(Y,\sigma,\eta)$. That the collusive value function is lower with a leniency program is due to two effects: i) $\phi_{NL}(Y,\eta) > \phi_\theta(Y,\eta)$ results in shorter cartel duration with a leniency program; and ii) when there is a leniency program, expected penalties upon cartel collapse are $\omega\gamma(Y-\alpha\mu)$ rather than $\sigma\gamma(Y-\alpha\mu)$, and the former are higher when $\omega > \sigma$. Third, $\psi_{NL}(Y,\sigma,\eta) > \psi_\theta(Y,\sigma,\eta)$ implies a weakly lower fixed point with a leniency program - $Y^*_{NL}(\sigma,\eta) \geq Y^*_\theta(\sigma,\eta)$ - and, therefore, a weakly lower equilibrium threshold: $\phi^*_{NL}(\sigma,\eta) \geq \phi^*_\theta(\sigma,\eta)$. This proves the cartel rate is no higher with a leniency program. If, in addition, there is a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) > \underline{\pi}$, then $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta)$ for a positive measure of values for η. From this result, one can then conclude that, holding σ fixed, the cartel rate is strictly lower with a leniency program.

Holding Y fixed, the threshold function for stable collusion is lower with a leniency program:

$$\begin{aligned}
& \phi_{NL}(Y,\sigma,\eta) - \phi_\theta(Y,\sigma,\eta) \\
= & \frac{\delta(1-\sigma)(1-\kappa)(Y-\alpha\mu)}{(\eta-1)[1-\delta(1-\kappa)]} \\
& - \left(\frac{\delta(1-\sigma)(1-\kappa)(Y-\alpha\mu) - [1-\delta(1-\kappa)][\sigma - \min\{\sigma,\theta\}]\gamma(Y-\alpha\mu)}{(\eta-1)[1-\delta(1-\kappa)]}\right) \\
= & \frac{(\sigma-\theta)\gamma(Y-\alpha\mu)}{\eta-1} > 0
\end{aligned} \tag{24}$$

because $\sigma > \theta$. Using $\phi_{NL}(Y,\eta) > \phi_\theta(Y,\eta)$,

$$\begin{aligned}
& \psi_{NL}(Y,\sigma,\eta) - \psi_\theta(Y,\sigma,\eta) \\
= & \int_{\underline{\pi}}^{\phi_{NL}(Y,\sigma,\eta)} \{(1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W] - (1-\delta)\sigma\gamma(Y-\alpha\mu)\} h(\pi)\,d\pi \\
& + \int_{\phi_{NL}(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W - (1-\delta)\sigma\gamma(Y-\alpha\mu)] h(\pi)\,d\pi \\
& - \int_{\underline{\pi}}^{\phi_\theta(Y,\sigma,\eta)} \{(1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W] - (1-\delta)\sigma\gamma(Y-\alpha\mu)\} h(\pi)\,d\pi \\
& - \int_{\phi_\theta(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W - \omega\gamma(Y-\alpha\mu)] h(\pi)\,d\pi
\end{aligned}$$

$$\begin{aligned}
= & \int_{\phi_\theta(Y,\sigma,\eta)}^{\phi_{NL}(Y,\sigma,\eta)} \{(1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W]\} h(\pi)\,d\pi - \int_{\phi_\theta(Y,\sigma,\eta)}^{\phi_{NL}(Y,\sigma,\eta)} (1-\delta)\sigma\gamma(Y-\alpha\mu) h(\pi)\,d\pi \\
& + \int_{\phi_{NL}(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W] h(\pi)\,d\pi - \int_{\phi_{NL}(Y,\sigma,\eta)}^{\overline{\pi}} (1-\delta)\sigma\gamma(Y-\alpha\mu) h(\pi)\,d\pi \\
& - \int_{\phi_\theta(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W] h(\pi)\,d\pi + \int_{\phi_\theta(Y,\sigma,\eta)}^{\overline{\pi}} \omega\gamma(Y-\alpha\mu) h(\pi)\,d\pi
\end{aligned}$$

$$= \int_{\phi_\theta(Y,\sigma,\eta)}^{\phi_{NL}(Y,\sigma,\eta)} \{(1-\delta)\pi + \delta[(1-\sigma)Y + \sigma W]\} h(\pi)\, d\pi - \int_{\phi_\theta(Y,\sigma,\eta)}^{\phi_{NL}(Y,\sigma,\eta)} [(1-\delta)\alpha\pi + \delta W] h(\pi)\, d\pi$$
$$- \int_{\phi_\theta(Y,\sigma,\eta)}^{\overline{\pi}} (1-\delta)\sigma\gamma(Y-\alpha\mu) h(\pi)\, d\pi + \int_{\phi_\theta(Y,\sigma,\eta)}^{\overline{\pi}} \omega\gamma(Y-\alpha\mu) h(\pi)\, d\pi$$

$$\begin{aligned} & \psi_{NL}(Y,\sigma,\eta) - \psi_\theta(Y,\sigma,\eta) \\ = & \int_{\phi_\theta(Y,\sigma,\eta)}^{\phi_{NL}(Y,\sigma,\eta)} [(1-\delta)(1-\alpha)\pi + \delta(1-\sigma)(Y-W)] h(\pi)\, d\pi \\ & + \int_{\phi_\theta(Y,\sigma,\eta)}^{\overline{\pi}} (1-\delta)(\omega-\sigma)\gamma(Y-\alpha\mu) h(\pi)\, d\pi. \end{aligned} \tag{25}$$

Given $\omega > \sigma$, (25) is non-negative. If $\underline{\pi} \geq \phi_{NL}(Y,\sigma,\eta)\,(> \phi_\theta(Y,\sigma,\eta))$ or $(\phi_{NL}(Y,\sigma,\eta) >)\, \phi_\theta(Y,\sigma,\eta) \geq \overline{\pi}$ then the first of the two terms in (25) is zero; otherwise, it is positive. If $\phi_\theta(Y,\sigma,\eta) \geq \overline{\pi}$ then the second term is zero; otherwise, it is positive.

Since it has just been shown that $\psi_{NL}(Y,\sigma,\eta) \geq \psi_\theta(Y,\sigma,\eta)$ then $Y^*_{NL}(\sigma,\eta) \geq Y^*_\theta(\sigma,\eta)$. Given

$$\phi^*(\sigma,\eta) \equiv \max\{\min\{\phi(Y^*(\sigma,\eta),\sigma,\eta),\overline{\pi}\},\underline{\pi}\},$$

it follows that $\phi^*_{NL}(\sigma,\eta) \geq \phi^*_\theta(\sigma,\eta)$.

Next we want to show: if there is positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) > \underline{\pi}$ then $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta)$ for a positive measure of values for η. If $\phi^*_{NL}(\sigma,\eta) < \overline{\pi}$ then either $\phi^*_{NL}(\sigma,\eta) > \underline{\pi}$ - so that $\phi^*_{NL}(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$ - or $\phi^*_{NL}(\sigma,\eta) = \underline{\pi}$; and if $\phi^*_{NL}(\sigma,\eta) > \underline{\pi}$ then either $\phi^*_{NL}(\sigma,\eta) < \overline{\pi}$ - so that $\phi^*_{NL}(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$ - or $\phi^*_{NL}(\sigma,\eta) = \overline{\pi}$. This results in two mutually exclusive cases: 1) there is a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$; and 2) there is not a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$ in which case there is a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) = \underline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}(\sigma,\eta) = \overline{\pi}$.

In considering case (1), first note that

$$\phi^*_{NL}(\sigma,\eta) = \phi(Y^*_{NL}(\sigma,\eta),\sigma,\eta) > \phi_\theta(Y^*_{NL}(\sigma,\eta),\sigma,\eta) \geq \phi_\theta(Y^*_\theta(\sigma,\eta),\sigma,\eta), \tag{26}$$

where the equality follows from $\phi^*_{NL}(\sigma,\eta) \in (\underline{\pi},\overline{\pi})$, the strict inequality follows from $\phi_{NL}(Y,\sigma,\eta) > \phi_\theta(Y,\sigma,\eta)$, and the weak inequality follows from $Y^*_{NL}(\sigma,\eta) \geq Y^*_\theta(\sigma,\eta)$. (26) implies $\overline{\pi} > \phi_\theta(Y^*_\theta(\sigma,\eta),\sigma,\eta)$ and, therefore,

$$\phi^*_\theta(\sigma,\eta) = \max\{\phi_\theta(Y^*_\theta(\sigma,\eta),\sigma,\eta),\underline{\pi}\}. \tag{27}$$

(26)-(27) allow us to conclude: $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta)$. Hence, for case (1), there is a positive measure of values for η for which $\phi^*_{NL}(\sigma,\eta) > \phi^*_\theta(\sigma,\eta)$. Under case (2), that $\phi^*_{NL}(\sigma,\eta)$ is weakly decreasing in (Lemma 2) implies $\exists\widehat{\eta}_{NL} \in (\underline{\eta},\overline{\eta})$ such that

$$\phi^*_{NL}(\sigma,\eta) = \begin{cases} \overline{\pi} & \text{if } \eta \in [\underline{\eta},\widehat{\eta}_{NL}] \\ \underline{\pi} & \text{if } \eta \in (\widehat{\eta}_{NL},\overline{\eta}] \end{cases}. \tag{28}$$

Note that, at the critical value $\widehat{\eta}_{NL}$,

$$\psi_{NL}\left(Y,\sigma,\widehat{\eta}_{NL}\right) \leq Y,\ \forall Y \in \left[\alpha\mu,\mu\right], \tag{29}$$

for suppose not. Then $\exists Y' \in (\alpha\mu,\mu)$ such that $\psi_{NL}\left(Y',\sigma,\widehat{\eta}_{NL}\right) > Y'$. By the continuity of ψ_{NL} in η, $\exists\xi > 0$ such that $\psi_{NL}\left(Y',\sigma,\widehat{\eta}_{NL}+\xi\right) > Y'$ which implies $Y^*_{NL}\left(\sigma,\widehat{\eta}_{NL}+\xi\right) > \alpha\mu$ and $\phi^*_{NL}\left(\sigma,\widehat{\eta}_{NL}+\xi\right) > \underline{\pi}$, but that contradicts (28). With (29) and $\psi_{NL}\left(Y,\sigma,\eta\right) > \psi_{\theta}\left(Y,\sigma,\eta\right)$, it follows $\exists\chi > 0$ such that $\psi_{\theta}\left(Y,\sigma,\widehat{\eta}_{NL}\right) < Y-\chi\ \forall Y \in [\alpha\mu,\mu]$ which implies, by the continuity of ψ_{θ} in η, $\exists\widehat{\eta}_{\theta} < \widehat{\eta}_{NL}$ such that $\phi^*_{\theta}\left(\sigma,\eta\right) = \underline{\pi}$ iff $\eta > \widehat{\eta}_{\theta}$. We then have that there is a positive measure of values of η - specifically, $\eta \in [\widehat{\eta}_{NL},\widehat{\eta}_{\theta})$ - for which

$$\phi^*_{NL}\left(\sigma,\eta\right) = \overline{\pi} > \underline{\pi} = \phi^*_{\theta}\left(\sigma,\eta\right).$$

This concludes the proof that: if there is positive measure of values for η such that $\phi^*_{NL}\left(\sigma,\eta\right) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}\left(\sigma,\eta\right) > \underline{\pi}$ then $\phi^*_{NL}\left(\sigma,\eta\right) > \phi^*_{\theta}\left(\sigma,\eta\right)$ for positive measure of values for η.

Whether with or without a leniency program, if the threshold for a type-η industry is $\widetilde{\phi}\left(\sigma,\eta\right)$ then the cartel rate is

$$\int_{\underline{\eta}}^{\overline{\eta}}\left[\frac{\kappa\left(1-\sigma\right)H\left(\widetilde{\phi}\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right)H\left(\widetilde{\phi}\left(\sigma,\eta\right)\right)}\right]g\left(\eta\right)d\eta. \tag{30}$$

Note that the cartel rate is increasing in $\widetilde{\phi}\left(\sigma,\eta\right)$. Given it has been shown $\phi^*_{NL}\left(\sigma,\eta\right) \geq \phi^*_{\theta}\left(\sigma,\eta\right)\ \forall\eta$, (30) implies $C_{NL}\left(\sigma\right) \geq C_{\theta}\left(\sigma\right)$. It has also been shown that: if there is a positive measure of values of η such that $\phi^*_{NL}\left(\sigma,\eta\right) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}\left(\sigma,\eta\right) > \underline{\pi}$ then there is a positive measure of values of η such that $\phi^*_{NL}\left(\sigma,\eta\right) > \phi^*_{\theta}\left(\sigma,\eta\right)$ and, therefore,

$$\frac{\kappa\left(1-\sigma\right)H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right)H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right)} > \frac{\kappa\left(1-\sigma\right)H\left(\phi^*_{\theta}\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right)H\left(\phi^*_{\theta}\left(\sigma,\eta\right)\right)}. \tag{31}$$

As (31) holds for a positive measure of values of η, (30) implies $C_{NL}\left(\sigma\right) > C_{\theta}\left(\sigma\right)$. ■

Proof of Theorem 7. When H is uniform, we will show that a type-η industry either is able to collude for all market conditions - $\phi^*_{NL}\left(\sigma,\eta\right) = \overline{\pi}$ - or is unable to collude for all market conditions - $\phi^*_{NL}\left(\sigma,\eta\right) = \underline{\pi}$. Hence, $\exists\widehat{\eta}_{NL}$ such that if $\eta \leq \widehat{\eta}_{NL}$ then collusion is always stable, and if $\eta > \widehat{\eta}_{NL}$ then collusion is never stable. A leniency program is shown to lower this threshold value and from that result it will be shown that introducing a leniency program raises the equilibrium value for σ and lowers the equilibrium cartel rate. The proof involves some tedious calculations associated with deriving the derivatives of $\psi_{NL}\left(Y\right)$ and solving for the threshold values; those calculations are available on request.

Under the assumptions that H is uniform and γ is sufficiently close to zero, if there is no leniency program then it can be shown that

$$\psi'_{NL}(Y) = \begin{cases} \frac{\delta\kappa}{1-\delta(1-\kappa)} - (1-\delta)\sigma\gamma & \text{if } Y \leq A \\ [(1-\delta)(1-\alpha)\phi_{NL}(Y) + \delta(1-\sigma)(Y-W)]\, h(\phi_{NL}(Y))\, \phi'_{NL}(Y) & \text{if } Y \in (A,B) \\ \quad + \left[\delta(1-\sigma) + \delta\sigma\frac{\partial W}{\partial Y}\right] H(\phi_{NL}(Y)) & \\ \quad + \delta\frac{\partial W}{\partial Y}\left[1 - H(\phi_{NL}(Y))\right] - (1-\delta)\sigma\gamma & \\ \delta\left(\frac{1-\delta(1-\kappa)-\sigma(1-\kappa)(1-\delta)}{1-\delta(1-\kappa)}\right) - (1-\delta)\sigma\gamma & \text{if } Y \geq B \end{cases}$$

where

$$\phi'_{NL}(Y) = \frac{\delta(1-\sigma)(1-\kappa)}{(\eta-1)[1-\delta(1-\kappa)]}$$

$$\begin{aligned} A &\equiv \alpha\mu + \frac{\underline{\pi}(\eta-1)[1-\delta(1-\kappa)]}{\delta(1-\sigma)(1-\kappa)} \\ B &\equiv \alpha\mu + \frac{\overline{\pi}(\eta-1)[1-\delta(1-\kappa)]}{\delta(1-\sigma)(1-\kappa)} \end{aligned}$$

When $Y \leq A$, $\phi_{NL}(Y) = \underline{\pi}$ so collusion is not stable for all market conditions; when $Y \in (A,B)$ then $\phi_{NL}(Y) \in (\underline{\pi}, \overline{\pi})$ so collusion is stable for some market conditions; and when $Y \geq B$ then $\phi_{NL}(Y) = \overline{\pi}$ so collusion is stable for all market conditions. Note that, when $\gamma \simeq 0$, $\psi'_{NL}(Y) \in (0,1)$ if $Y \leq A$ or $Y \geq B$. $\psi_{NL}(Y)$ is linear when Y does not affect $\phi_{NL}(Y)$ and is quadratic when it does:

$$\psi''_{NL}(Y) = \begin{cases} 0 & \text{if } Y \leq A \\ \left[(1-\delta)(1-\alpha)\phi'_{NL}(Y) + 2\delta(1-\sigma)\left(1 - \frac{\partial W}{\partial Y}\right)\right] h(\phi_{NL}(Y))\, \phi'_{NL}(Y) & \text{if } Y \in (A,B) \\ 0 & \text{if } Y \geq B \end{cases}$$

$$\psi'''_{NL}(Y) = 0, \ \forall Y.$$

If $A \geq \mu$ then ψ_{NL} is linear and $\psi'_{NL}(Y) \in (0,1)$, $\forall Y \in [\alpha\mu, \mu]$; see Figure 3. Given that $\psi_{NL}(\alpha\mu) = \alpha\mu$, it follows that $\psi_{NL}(Y) < Y \ \forall Y > \alpha\mu$ which implies that $\alpha\mu$ is the unique fixed point of $\alpha\mu$; this industry type never colludes. If $B \geq \mu$ then again there is a unique fixed point of $\alpha\mu$ by the following argument. When $B \geq \mu$, ψ_{NL} is weakly convex $\forall Y \in [\alpha\mu, \mu]$ (it is linear then strictly convex); see Figure 4. Given that $\psi_{NL}(\alpha\mu) = \alpha\mu$ and $\psi_{NL}(\mu) < \mu$ then $\psi_{NL}(Y) < Y \forall Y \in (\alpha\mu, \mu]$. Thus, a necessary condition for there to be a fixed point exceeding $\alpha\mu$ is that $B < \mu$. Note that B is increasing in η and $\lim_{\eta \to 1} B = \alpha\mu$. Thus, if η is sufficiently low then $B < \mu$. Values of η such that $B < \mu$ exist by supposition because it was presumed that the equilibrium cartel rate without a leniency program is positive which means some industry types are able to collude.

Insert Figure 3 here

Insert Figure 4 here

Assume $B < \mu$ and $\psi_{NL}(B) < B$; see Figure 5. Given $\psi_{NL}(\alpha\mu) = \alpha\mu$ and ψ_{NL} is weakly convex $\forall Y \in [\alpha\mu, B]$, $\psi_{NL}(B) < B$ implies $\psi_{NL}(Y) < Y$ $\forall Y \in (\alpha\mu, B)$. Next note that for $Y \in [B, \mu]$, ψ_{NL} is linear and $\psi'_{NL}(Y) \in (0, 1)$ in which case $\psi_{NL}(B) < B$ implies $\psi_{NL}(Y) < Y$ $\forall Y \in [B, \mu]$. We then have that $\psi_{NL}(B) < B$ implies $\psi_{NL}(Y) < Y$ $\forall Y \in (\alpha\mu, \mu]$ which means there is a unique fixed point of $\alpha\mu$.

Insert Figure 5 here

Finally, assume $B < \mu$ and $\psi_{NL}(B) \geq B$; see Figure 6. If $\psi_{NL}(B) > B$ then there is a unique fixed point in $(\alpha\mu, B)$ where uniqueness comes from ψ_{NL} being weakly convex over $[\alpha\mu, B]$. Since ψ_{NL} is linear, $\psi'_{NL}(Y) \in (0, 1)$ for $Y \in [B, \mu]$, and $\psi_{NL}(\mu) < \mu$ then there is a second fixed point in (B, μ). Thus, if $\psi_{NL}(B) > B$ then there are two fixed points exceeding $\alpha\mu$. When $\psi_{NL}(B) = B$, there is one fixed point exceeding $\alpha\mu$ which is B. In sum, when $\psi_{NL}(B) \geq B$, the maximal fixed point is at least B which means that it occurs where $\phi_{NL}(Y) = \overline{\pi}$. Using the expression for $\psi_{NL}(Y)$ when $\phi_{NL}(Y) = \overline{\pi}$ (which is the LHS of (32)), the collusive value $Y^*_{NL}(\sigma, \eta)$ is the unique solution to:

$$(1-\delta)\mu + \delta Y^*_{NL}(\sigma,\eta) - \left[\delta\sigma\left(\frac{(1-\kappa)(1-\delta)}{1-\delta(1-\kappa)}\right) + (1-\delta)\sigma\gamma\right](Y^*_{NL}(\sigma,\eta) - \alpha\mu) = Y^*_{NL}(\sigma,\eta). \tag{32}$$

Insert Figure 6 here

Summarizing, if η is such that $B \geq \mu$ or $B < \mu$ and $\psi_{NL}(B) < B$ then collusion is unstable for all market conditions so $\phi^*_{NL}(\sigma, \eta) = \underline{\pi}$. If $B < \mu$ and $\psi_{NL}(B) \geq B$ then collusion is stable for all market conditions so $\phi^*_{NL}(\sigma, \eta) = \overline{\pi}$. Hence, $\widehat{\eta}_{NL}(\sigma)$ is the lowest value for η such that $\phi^*_{NL}(\sigma, \eta) = \overline{\pi}$. Given ψ_{NL} is continuous, $\widehat{\eta}_{NL}(\sigma)$ is defined by $\psi_{NL}(B, \sigma, \widehat{\eta}_{NL}) = B$, which takes the explicit form:

$$(1-\delta)\mu + \delta B - \left[\delta\sigma\left(\frac{(1-\kappa)(1-\delta)}{1-\delta(1-\kappa)}\right) + (1-\delta)\sigma\gamma\right](B - \alpha\mu) = B.$$

Substituting for B and solving for $\widehat{\eta}_{NL}(\sigma)$ (derivations are in the appendix),

$$\widehat{\eta}_{NL}(\sigma) = 1 + \left(\frac{\mu}{\overline{\pi}}\right)\left(\frac{\delta(1-\sigma)(1-\kappa)(1-\alpha)}{\delta\sigma(1-\kappa) + (1+\sigma\gamma)(1-\delta(1-\kappa))}\right) \tag{33}$$

Without a leniency program, cartels then emerge only in industries for which $\eta \leq \widehat{\eta}_{NL}(\sigma)$ and those cartels never internally collapse though are shut down by the CA at a rate of σ per period.

We now need to repeat the analysis for when there is a full leniency program ($\theta = 0$). The steps are exactly the same as above; it is just that some expressions are different.

$$\psi_L'(Y) = \begin{cases} \frac{\delta\kappa}{1-\delta(1-\kappa)} - (1-\delta)\,\omega\gamma & \text{if } Y \leq C \\ [(1-\delta)(1-\alpha)\,\phi_L(Y) + \delta(1-\sigma)(Y-W) - (1-\delta)(\sigma-\omega)\,\gamma\,(Y-\alpha\mu)] \times & \text{if } Y \in (C,D) \\ h(\phi_L(Y))\,\phi_L'(Y) + \left[\delta(1-\sigma) + \delta\sigma\frac{\partial W}{\partial Y} - (1-\delta)\,\sigma\gamma\right] H(\phi_L(Y)) & \\ + \left[\delta\frac{\partial W}{\partial Y} - (1-\delta)\,\omega\gamma\right](1 - H(\phi_L(Y))) & \\ \delta\left(1 - \sigma\left(\frac{(1-\kappa)(1-\delta)}{1-\delta(1-\kappa)}\right)\right) - (1-\delta)\,\sigma\gamma & \text{if } Y \geq D \end{cases}$$

where

$$\phi_L'(Y) = \frac{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)](\sigma-\theta)\,\gamma}{(\eta-1)[1-\delta(1-\kappa)]}$$

$$\begin{aligned} C &\equiv \alpha\mu + \frac{\underline{\pi}(\eta-1)[1-\delta(1-\kappa)]}{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)](\sigma-\theta)\,\gamma} \\ D &\equiv \alpha\mu + \frac{\overline{\pi}(\eta-1)[1-\delta(1-\kappa)]}{\delta(1-\sigma)(1-\kappa) - [1-\delta(1-\kappa)](\sigma-\theta)\,\gamma} \end{aligned}$$

Furthermore, $\psi_L'(Y) \in (0,1)$ if $Y \leq C$ or $Y \geq D$.

$$\psi_L''(Y) = \begin{cases} 0 & \text{if } Y \leq C \\ \left[(1-\delta)(1-\alpha)\,\phi_L'(Y) + 2\delta(1-\sigma)\left(1 - \frac{\partial W}{\partial Y}\right) - 2(1-\delta)(\sigma-\omega)\,\gamma\right] \times & \text{if } Y \in (C,D) \\ h(\phi_L(Y))\,\phi_L'(Y) & \\ 0 & \text{if } Y \geq D \end{cases}$$

$$\psi_L'''(Y) = 0,\ \forall Y.$$

Again, ψ_L is a linear then quadratic then linear function.

If $C \geq \mu$ then there is a unique fixed point of $\alpha\mu$, in which case this industry type never cartelizes. If $D \geq \mu$ then again there is a unique fixed point of $\alpha\mu$. Thus, a necessary condition for there to be a fixed point exceeding $\alpha\mu$ is that $D < \mu$. Assume η is sufficiently low so that $D < \mu$. If no values for η in $\left[\underline{\eta}, \overline{\eta}\right]$ exist whereby $D < \mu$, it means that all industry types cannot collude for all market conditions; hence, the cartel rate with a leniency program is zero, which proves that it is lower than when there is no leniency program. Thus, let us suppose that there are values for η such that $D < \mu$.

If $\psi_L(D) < D$ then, as argued for the case of no leniency program, there are no fixed points exceeding $\alpha\mu$. If $\psi_L(D) > (=)\,D$ then there are two (one) fixed points exceeding $\alpha\mu$ and the maximal fixed point is at least D which means that it occurs where $\phi_L(Y) = \overline{\pi}$. Thus, if η is such that $D \geq \mu$ or $D < \mu$ and $\psi_L(D) < D$ then collusion is unstable for all market conditions so $\phi_L^*(\sigma,\eta) = \underline{\pi}$. If $D < \mu$ and $\psi_L(D) \geq D$ then collusion is stable for all market conditions so $\phi_L^*(\sigma,\eta) = \overline{\pi}$. $\widehat{\eta}_L(\sigma)$ is then the highest value for η such that $\psi_L(D) \geq D$ which, by the continuity of ψ_L, is that value for η such that $\psi_L(D) = D$. Given $\phi_L(Y) = \overline{\pi}$, $\psi_L(D) = D$ takes the form:

$$(1-\delta)\mu + \delta D - \delta\sigma\left(\frac{(1-\kappa)(1-\delta)(D-\alpha\mu)}{1-\delta(1-\kappa)}\right) - (1-\delta)\sigma\gamma(D-\alpha\mu) = D.$$

Substituting for D and solving for $\widehat{\eta}_L(\sigma)$ yields:

$$\widehat{\eta}_L(\sigma) = 1 + \left(\frac{\mu}{\overline{\pi}}\right)\left(\frac{(1-\alpha)(\delta(1-\sigma)(1-\kappa) - (1-\delta(1-\kappa))\sigma\gamma)}{\delta\sigma(1-\kappa) + (1+\sigma\gamma)(1-\delta(1-\kappa))}\right) \tag{34}$$

Let us now combine the analyses for the cases of no leniency program and a full leniency program. For when there is no leniency program, a type-η cartel colludes for all market conditions when $\eta \leq \widehat{\eta}_{NL}(\sigma)$, and does not collude for all market conditions when $\eta > \widehat{\eta}_{NL}(\sigma)$. For when there is a full leniency program, a type-η cartel colludes for all market conditions when $\eta \leq \widehat{\eta}_L(\sigma)$, and does not collude for all market conditions when $\eta > \widehat{\eta}_L(\sigma)$. Using the expressions in (33) and (34),

$$\widehat{\eta}_{NL}(\sigma) - \widehat{\eta}_L(\sigma) = \left(\frac{\mu}{\overline{\pi}}\right)\left(\frac{(1-\delta(1-\kappa))\sigma\gamma}{\delta\sigma(1-\kappa) + (1+\sigma\gamma)(1-\delta(1-\kappa))}\right) > 0.$$

Hence, holding σ fixed, fewer industry types cartelize when there is a leniency program.

Whether or not there is a leniency program, the formula for the cartel rate is

$$\int_{\underline{\eta}}^{\overline{\eta}} \left[\frac{\kappa(1-\sigma)H\left(\widetilde{\phi}(\sigma,\eta)\right)}{1-(1-\kappa)(1-\sigma)H\left(\widetilde{\phi}(\sigma,\eta)\right)}\right] g(\eta)\,d\eta,$$

where $\widetilde{\phi}(\sigma,\eta) = \phi_{NL}^*(\sigma,\eta)$ without a leniency program, and $\widetilde{\phi}(\sigma,\eta) = \phi_L^*(\sigma,\eta)$ with a leniency program. Given that

$$\phi_{NL}^*(\sigma,\eta) = \begin{cases} \overline{\pi} & \text{if } \eta \leq \widehat{\eta}_{NL}(\sigma) \\ \underline{\pi} & \text{if } \eta > \widehat{\eta}_{NL}(\sigma) \end{cases}$$

then the cartel rate function without a leniency program is

$$C_{NL}(\sigma) = \int_{\underline{\eta}}^{\widehat{\eta}_{NL}(\sigma)} \left[\frac{\kappa(1-\sigma)}{1-(1-\kappa)(1-\sigma)}\right] g(\eta)\,d\eta = \frac{\kappa(1-\sigma)G(\widehat{\eta}_{NL}(\sigma))}{1-(1-\kappa)(1-\sigma)}.$$

Given that

$$\phi_L^*(\sigma,\eta) = \begin{cases} \overline{\pi} & \text{if } \eta \leq \widehat{\eta}_L(\sigma) \\ \underline{\pi} & \text{if } \eta > \widehat{\eta}_L(\sigma) \end{cases}$$

then the cartel rate function with a leniency program is

$$C_L(\sigma) = \int_{\underline{\eta}}^{\widehat{\eta}_L(\sigma)} \left[\frac{\kappa(1-\sigma)}{1-(1-\kappa)(1-\sigma)}\right] g(\eta)\, d\eta = \frac{\kappa(1-\sigma) G(\widehat{\eta}_L(\sigma))}{1-(1-\kappa)(1-\sigma)}.$$

Thus, holding σ fixed, $\widehat{\eta}_{NL}(\sigma) > \widehat{\eta}_L(\sigma)$ implies that a leniency program reduces the cartel rate:

$$C_{NL}(\sigma) = \frac{\kappa(1-\sigma) G(\widehat{\eta}_{NL}(\sigma))}{1-(1-\kappa)(1-\sigma)} > \frac{\kappa(1-\sigma) G(\widehat{\eta}_L(\sigma))}{1-(1-\kappa)(1-\sigma)} = C_L(\sigma). \tag{35}$$

The final step is to endogenize σ. Without a leniency program, σ_{NL}^* is defined by:

$$\sigma_{NL}^* = qrp\left(\frac{qr\kappa(1-\sigma_{NL}^*) G(\widehat{\eta}_{NL}(\sigma_{NL}^*))}{1-(1-\kappa)(1-\sigma_{NL}^*)}\right),$$

and, given that it is the maximal fixed point,

$$\sigma \gtreqless qrp\left(\frac{qr\kappa(1-\sigma) G(\widehat{\eta}_{NL}(\sigma))}{1-(1-\kappa)(1-\sigma)}\right) \text{ as } \sigma \gtreqless \sigma_{NL}^*. \tag{36}$$

Since cartels never collapse - they either form and are always stable or do not form - then there are no leniency applications (recall that only dying cartels apply for leniency). As a result, the equilibrium cartel rate with a leniency program σ_L^* is

$$\sigma_L^* = qrp\left(\frac{qr\kappa(1-\sigma_L^*) G(\widehat{\eta}_L(\sigma_L^*))}{1-(1-\kappa)(1-\sigma_L^*)}\right). \tag{37}$$

Given p is decreasing, it follows from (35):

$$qrp\left(\frac{qr\kappa(1-\sigma) G(\widehat{\eta}_L(\sigma))}{1-(1-\kappa)(1-\sigma)}\right) > qrp\left(\frac{qr\kappa(1-\sigma) G(\widehat{\eta}_{NL}(\sigma))}{1-(1-\kappa)(1-\sigma)}\right). \tag{38}$$

Using (36), (38) implies

$$qrp\left(\frac{qr\kappa(1-\sigma_{NL}^*) G(\widehat{\eta}_L(\sigma_{NL}^*))}{1-(1-\kappa)(1-\sigma_{NL}^*)}\right) > \sigma_{NL}^*, \tag{39}$$

and, therefore, $\sigma_L^* > \sigma_{NL}^*$. Given $C_{NL}(\sigma) > C_L(\sigma)$ from (35), and $C_L(\sigma)$ is decreasing in σ (Lemma 4), $\sigma_L^* > \sigma_{NL}^*$ implies the equilibrium cartel rate is lower with a leniency program:

$$C_{NL}(\sigma_{NL}^*) > C_L(\sigma_{NL}^*) > C_L(\sigma_L^*).$$

■

Proof of Theorem 8. Given $\sigma^* \in (0, \omega)$ and $\theta = 0$, by Theorem 6 we have that $C_{NL}(\sigma) \geq C_L(\sigma)$ and, when there is positive measure of values for η such that $\phi_{NL}^*(\sigma, \eta) < \overline{\pi}$ and a positive measure of values for η such that $\phi_{NL}^*(\sigma, \eta) > \underline{\pi}$, $C_{NL}(\sigma) > C_L(\sigma)$. To prove this theorem, it is then sufficient to show $\sigma_L^* > \sigma_{NL}^*$.

σ^*_{NL} and σ^*_L are defined by:

$$\sigma^*_{NL} = qrp\left(qr\int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma^*_{NL},\eta\right) g\left(\eta\right) d\eta\right)$$

where

$$C_{NL}\left(\sigma,\eta\right) = \frac{\kappa\left(1-\sigma\right) H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right) H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right)},$$

and

$$\begin{aligned}\sigma^*_L &= p\left(\lambda\int_{\underline{\eta}}^{\overline{\eta}} \left(1-H\left(\phi^*_L\left(\sigma^*_L,\eta\right)\right)\right) C_L\left(\sigma^*_L,\eta\right) g\left(\eta\right) d\eta\right. \\ &\qquad \left. +qr\int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*_L\left(\sigma^*_L,\eta\right)\right) C_L\left(\sigma^*_L,\eta\right) g\left(\eta\right) d\eta\right)\end{aligned}$$

where

$$C_L\left(\sigma,\eta\right) = \frac{\kappa\left(1-\sigma\right) H\left(\phi^*_L\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right) H\left(\phi^*_L\left(\sigma,\eta\right)\right)}.$$

If $\phi^*_{NL}\left(\sigma,\eta\right) > (=)\, \phi^*_L\left(\sigma,\eta\right)$ then

$$H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right) > (=)\, H\left(\phi^*_L\left(\sigma,\eta\right)\right)$$

and

$$C_{NL}\left(\sigma,\eta\right) > (=)\, C_L\left(\sigma,\eta\right),$$

in which case,

$$H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right) C_{NL}\left(\sigma,\eta\right) > (=)\, H\left(\phi^*_L\left(\sigma,\eta\right)\right) C_L\left(\sigma,\eta\right).$$

It is immediate that if

$$H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right) C_{NL}\left(\sigma,\eta\right) \geq H\left(\phi^*_L\left(\sigma,\eta\right)\right) C_L\left(\sigma,\eta\right), \forall\eta \tag{40}$$

and

$$H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right) C_{NL}\left(\sigma,\eta\right) > H\left(\phi^*_L\left(\sigma,\eta\right)\right) C_L\left(\sigma,\eta\right), \text{ for positive measure of } \eta \tag{41}$$

then

$$\int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*_{NL}\left(\sigma,\eta\right)\right) C_{NL}\left(\sigma,\eta\right) g\left(\eta\right) d\eta > \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*_L\left(\sigma,\eta\right)\right) C_L\left(\sigma,\eta\right) g\left(\eta\right) d\eta. \tag{42}$$

(40) is always true and (41) is true when there is positive measure of values for η such that $\phi^*_{NL}\left(\sigma,\eta\right) < \overline{\pi}$ and a positive measure of values for η such that $\phi^*_{NL}\left(\sigma,\eta\right) > \underline{\pi}$.

Evaluate (42) at $\sigma = \sigma^*_{NL}$:

$$\int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*_{NL}\left(\sigma^*_{NL},\eta\right)\right) C_{NL}\left(\sigma^*_{NL},\eta\right) g\left(\eta\right) d\eta > \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi^*_L\left(\sigma^*_{NL},\eta\right)\right) C_L\left(\sigma^*_{NL},\eta\right) g\left(\eta\right) d\eta. \tag{43}$$

Noting that σ_{NL}^* does not depend on λ, if λ is sufficiently small then it follows from (43):

$$\begin{aligned} & qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_{NL}^*\left(\sigma_{NL}^*, \eta\right)\right) C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \\ > \;& \lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_{NL}^*, \eta\right)\right)\right) C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta + \\ & qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_{NL}^*, \eta\right)\right) C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \end{aligned} \tag{44}$$

Given that p is decreasing then (44) implies (when λ is sufficiently small):

$$\begin{aligned} & qrp \left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_{NL}^*, \eta\right)\right)\right) C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \right. \\ & \left. + qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_{NL}^*, \eta\right)\right) C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \right) \\ > \;& qrp \left(qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_{NL}^*\left(\sigma_{NL}^*, \eta\right)\right) C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \right) \\ > \;& qrp \left(qr \int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \right) = \sigma_{NL}^*. \end{aligned}$$

Hence,

$$\begin{aligned} & qrp \left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma_{NL}^*, \eta\right)\right)\right) C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \right. \\ & \left. + qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma_{NL}^*, \eta\right)\right) C_L\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta \right) \\ > \;& \sigma_{NL}^* \end{aligned}$$

and thus $\sigma_L^* > \sigma_{NL}^*$.

In proving $\sigma_L^* > \sigma_{NL}^*$, the preceding analysis presumed $\omega > \sigma$. If, contrary to that presumption, $\sigma_L^* \geq \omega$ then the supposition that $\omega > \sigma_{NL}^*$ would again imply $\sigma_L^* > \sigma_{NL}^*$. ■

Proof of Theorem 9. Given that $\sigma = qrs$ and r,$s \in [0, 1]$ (hence, are bounded), it is immediate that

$$\lim_{q \to 0} \sigma_{NL}^* = 0, \lim_{q \to 0} \sigma_L^* = 0,$$

which implies

$$\lim_{q \to 0} C_{NL}\left(\sigma_{NL}^*\right) = \lim_{\sigma \to 0} C_{NL}\left(\sigma\right), \;\; \lim_{q \to 0} C_L\left(\sigma_L^*\right) = \lim_{\sigma \to 0} C_L\left(\sigma\right).$$

To show the equilibrium cartel rate is lower with a leniency program, it is then sufficient to prove:

$$\lim_{\sigma \to 0} C_{NL}(\sigma) > \lim_{\sigma \to 0} C_L(\sigma). \tag{45}$$

Given $\theta = 0 < \omega$ then $\sigma \in (\theta, \omega)$ holds as $\sigma \to 0$ in which case Theorem 6 proves (45). ■

Proof of Theorem 10. The first step is to show that, as the penalty multiple γ goes to zero, the cartel rate function is the same with and without a leniency program:

$$\lim_{\gamma \to 0} C_{NL}(\sigma) = \lim_{\gamma \to 0} C_L(\sigma), \ \forall \sigma.$$

The second step is to show that, as $\gamma \to 0$, non-leniency enforcement is weaker with a leniency program:

$$\lim_{\gamma \to 0} \sigma^*_{NL} > \lim_{\gamma \to 0} \sigma^*_L.$$

These two results together imply that the equilibrium cartel rate with a leniency program is higher than without a leniency program when $\gamma \simeq 0$.

For the first step, let us begin by considering the thresholds for stable collusion. Without a leniency program,

$$\phi_{NL}(Y, \sigma, \eta) = \frac{\delta(1-\sigma)(1-\kappa)(Y - \alpha\mu)}{(\eta - 1)[1 - \delta(1-\kappa)]}$$

and, trivially,[26]

$$\lim_{\gamma \to 0} \phi_{NL}(Y, \sigma, \eta) = \frac{\delta(1-\sigma)(1-\kappa)(Y - \alpha\mu)}{(\eta - 1)[1 - \delta(1-\kappa)]}.$$

With a full leniency program,

$$\phi_L(Y, \sigma, \eta) = \frac{\delta(1-\sigma)(1-\kappa)(Y - \alpha\mu) - [1 - \delta(1-\kappa)]\sigma\gamma(Y - \alpha\mu)}{(\eta - 1)[1 - \delta(1-\kappa)]}$$

and

$$\lim_{\gamma \to 0} \phi_L(Y, \sigma, \eta) = \frac{\delta(1-\sigma)(1-\kappa)(Y - \alpha\mu)}{(\eta - 1)[1 - \delta(1-\kappa)]}.$$

Hence,

$$\lim_{\gamma \to 0} \phi_{NL}(Y, \sigma, \eta) = \lim_{\gamma \to 0} \phi_L(Y, \sigma, \eta). \tag{46}$$

Turning to the collusive value functions, we have without a leniency program:

$$\begin{aligned} \psi_{NL}(Y, \sigma, \eta) &= \int_{\underline{\pi}}^{\phi_{NL}(Y,\sigma,\eta)} [(1-\delta)\pi + \delta(1-\sigma)Y + \delta\sigma W] h(\pi)\, d\pi \\ &\quad + \int_{\phi_{NL}(Y,\sigma,\eta)}^{\overline{\pi}} [(1-\delta)\alpha\pi + \delta W] h(\pi)\, d\pi - (1-\delta)\sigma\gamma(Y - \alpha\mu), \end{aligned}$$

[26] Recall that $\phi_{NL}(Y, \sigma, \eta)$ comes out of the ICC and is the market condition that makes a firm indifferent between colluding and cheating. That γ does not matter is because the expected penalty is the same whether a firm sets the collusive price or cheats and undercuts the collusive price set by the other firms.

and with a full leniency program:

$$\psi_L(Y,\sigma,\eta) = \int_{\underline{\pi}}^{\phi_L(Y,\sigma,\eta)} \left[(1-\delta)\pi + \delta(1-\sigma)Y + \delta\sigma W - (1-\delta)\sigma\gamma(Y-\alpha\mu)\right] h(\pi)\, d\pi + \int_{\phi_L(Y,\sigma,\eta)}^{\overline{\pi}} \left[(1-\delta)\alpha\pi + \delta W - (1-\delta)\omega\gamma(Y-\alpha\mu)\right] h(\pi)\, d\pi.$$

Using (46),

$$\begin{aligned} \lim_{\gamma\to 0}\psi_{NL}(Y,\sigma,\eta) &= \lim_{\gamma\to 0}\psi_L(Y,\sigma,\eta) \qquad (47)\\ &= \int_{\underline{\pi}}^{\phi_{NL}(Y,\sigma,\eta)} \left[(1-\delta)\pi + \delta(1-\sigma)Y + \delta\sigma W\right] h(\pi)\, d\pi \\ &\quad + \int_{\phi_{NL}(Y,\sigma,\eta)}^{\overline{\pi}} \left[(1-\delta)\alpha\pi + \delta W\right] h(\pi)\, d\pi. \end{aligned}$$

Generically, (47) implies

$$\lim_{\gamma\to 0} Y_{NL}^*(\sigma,\eta) = \lim_{\gamma\to 0} Y_L^*(\sigma,\eta). \qquad (48)$$

It follows from (46) and (47) that:

$$\lim_{\gamma\to 0} \phi_{NL}^*(\sigma,\eta) = \lim_{\gamma\to 0} \phi_L^*(\sigma,\eta). \qquad (49)$$

Given σ, the cartel rate without and with a leniency program, respectively, is:

$$C_{NL}(\sigma) = \int_{\underline{\eta}}^{\overline{\eta}} C_{NL}(\sigma,\eta)\, g(\eta)\, d\eta = \int_{\underline{\eta}}^{\overline{\eta}} \left[\frac{\kappa(1-\sigma)H(\phi_{NL}^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi_{NL}^*(\sigma,\eta))}\right] g(\eta)\, d\eta$$

$$C_L(\sigma) = \int_{\underline{\eta}}^{\overline{\eta}} C_L(\sigma,\eta)\, g(\eta)\, d\eta = \int_{\underline{\eta}}^{\overline{\eta}} \left[\frac{\kappa(1-\sigma)H(\phi_L^*(\sigma,\eta))}{1-(1-\kappa)(1-\sigma)H(\phi_L^*(\sigma,\eta))}\right] g(\eta)\, d\eta.$$

Using (49),

$$\lim_{\gamma\to 0} C_{NL}(\sigma) = \lim_{\gamma\to 0} C_L(\sigma). \qquad (50)$$

To prove the second step, we want to first show that, when $\lambda \simeq 1$ and $\gamma \simeq 0$,

$$\begin{aligned} & p\left(qr\int_{\underline{\eta}}^{\overline{\eta}} C_{NL}(\sigma,\eta)\, g(\eta)\, d\eta\right) \qquad (51)\\ > \;& p\left(\lambda\int_{\underline{\eta}}^{\overline{\eta}} (1-H(\phi_L^*(\sigma,\eta)))\, C_L(\sigma,\eta)\, g(\eta)\, d\eta + qr\int_{\underline{\eta}}^{\overline{\eta}} H(\phi_L^*(\sigma,\eta))\, C_L(\sigma,\eta)\, g(\eta)\, d\eta\right). \end{aligned}$$

Given p is strictly decreasing, (51) holds iff

$$\begin{aligned} & \lambda\int_{\underline{\eta}}^{\overline{\eta}} (1-H(\phi_L^*(\sigma,\eta)))\, C_L(\sigma,\eta)\, g(\eta)\, d\eta + qr\int_{\underline{\eta}}^{\overline{\eta}} H(\phi_L^*(\sigma,\eta))\, C_L(\sigma,\eta)\, g(\eta)\, d\eta \\ > \;& qr\int_{\underline{\eta}}^{\overline{\eta}} C_{NL}(\sigma,\eta)\, g(\eta)\, d\eta \end{aligned}$$

or, equivalently,

$$\begin{aligned} & \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma,\eta\right)\right)\right)\left[\lambda C_L\left(\sigma,\eta\right) - qrC_{NL}\left(\sigma,\eta\right)\right] g\left(\eta\right) d\eta \\ > \quad & qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma,\eta\right)\right)\left[C_{NL}\left(\sigma,\eta\right) - C_L\left(\sigma,\eta\right)\right] g\left(\eta\right) d\eta. \end{aligned} \tag{52}$$

Given (50), (52) holds as $\gamma \to 0$ iff

$$(\lambda - qr) \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_{NL}^*\left(\sigma,\eta\right)\right)\right) C_{NL}\left(\sigma,\eta\right) g\left(\eta\right) d\eta > 0. \tag{53}$$

By (12),

$$\int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_{NL}^*\left(\sigma,\eta\right)\right)\right) C_{NL}\left(\sigma,\eta\right) g\left(\eta\right) d\eta > 0 \tag{54}$$

holds for $\sigma = \sigma_{NL}^*$. Given $qr < 1$ then $\lambda \simeq 1$ implies $\lambda > qr$ and (53) holds. We have then shown that there exists $\widehat{\lambda} < 1$ and $\widehat{\gamma} > 0$ such that if $(\gamma, \lambda) \in [0, \widehat{\gamma}] \times \left[\widehat{\lambda}, 1\right]$ then (51) holds, generically, in a small neighborhood of $\sigma = \sigma_{NL}^*$.

For when there is no leniency program, σ_{NL}^* is defined by:

$$\sigma_{NL}^* = qrp\left(qr \int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta\right).$$

As it is the maximal fixed point then:

$$\sigma - qrp\left(qr \int_{\underline{\eta}}^{\overline{\eta}} C_{NL}\left(\sigma,\eta\right) g\left(\eta\right) d\eta\right) \gtreqqless 0 \text{ as } \sigma \gtreqqless \sigma_{NL}^*. \tag{55}$$

Hence, using (51), it follows from (55) that there exists $\widehat{\lambda} < 1$ and $\widehat{\gamma} > 0$ such that if $(\gamma, \lambda) \in [0, \widehat{\gamma}] \times \left[\widehat{\lambda}, 1\right]$ then $\exists \varepsilon > 0$ such that

$$\begin{aligned} & \sigma - qrp\left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma,\eta\right)\right)\right) C_L\left(\sigma,\eta\right) g\left(\eta\right) d\eta \right. \\ & \left. + qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma,\eta\right)\right) C_L\left(\sigma,\eta\right) g\left(\eta\right) d\eta\right) \\ > \quad & 0, \forall \sigma \geq \sigma_{NL}^* - \varepsilon. \end{aligned} \tag{56}$$

Given the continuity of

$$p\left(\lambda \int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_L^*\left(\sigma,\eta\right)\right)\right) C_L\left(\sigma,\eta\right) g\left(\eta\right) d\eta + qr \int_{\underline{\eta}}^{\overline{\eta}} H\left(\phi_L^*\left(\sigma,\eta\right)\right) C_L\left(\sigma,\eta\right) g\left(\eta\right) d\eta\right)$$

in σ (see the proof of Theorem 5), (56) implies the maximal fixed point σ_L^* is less than $\sigma_{NL}^* - \varepsilon$. Given (50) and having just shown

$$\lim_{\gamma \to 0} \sigma_{NL}^* > \lim_{\gamma \to 0} \sigma_L^*,$$

it follows that

$$\lim_{\gamma \to 0} C_L\left(\sigma_L^*\right) > \lim_{\gamma \to 0} C_{NL}\left(\sigma_{NL}^*\right).$$

■

References

[1] Apesteguia, Jose, Martin Dufwenberg, and Reinhard Selten, "Blowing the Whistle," *Economic Theory*, 31 (2007), 143-166.

[2] Aubert, Cecile, Patrick Rey, and William Kovacic, 2003, "The Impact of Leniency and Whistle-Blowing Programs on Cartels," *International Journal of Industrial Organization*, 24 (2006), 1241-1266.

[3] Bigoni, Maria, Sven-Olof Fridolfsson, Chloé Le Coq, and Giancarlo Spagnolo, "Fines, Leniency, and Rewards in Antitrust," *RAND Journal of Economics*, 43 (2012), 368-390.

[4] Borrell, Joan-Ramon, Juan Luis Jiménez, and Carmen García, "Evaluating Antitrust Leniency Programs," Universitat de Barcelona, January 2012.

[5] Brenner, Steffen, "An Empirical Study of the European Corporate Leniency Program," *International Journal of Industrial Organization,* 27 (2009), 639-645.

[6] Chen, Joe and Joseph E. Harrington, Jr., "The Impact of the Corporate Leniency Program on Cartel Formation and the Cartel Price Path," in *The Political Economy of Antitrust*, V. Ghosal and J. Stennek, eds., Elsevier, 2007.

[7] Chen, Zhijun and Patrick Rey, "On the Design of Leniency Programs," University of Toulouse, 2007.

[8] Choi, Jay Pil and Heiko Gerlach, "Global Cartels, Leniency Programs and International Antitrust Cooperation," Michigan State University, March 2010.

[9] Connor, John M., "Cartel Amnesties Granted: Worldwide Whistleblowers," Purdue University, 2008.

[10] Dijkstra, Peter T., Marco A. Haan, and Lambert Schoonbeek, "Leniency Programs and the Design of Antitrust: Experimental Evidence with Unrestricted Communication," University of Groningen, 2011.

[11] Ellis, Christopher J. and Wesley W. Wilson, "Cartels, Price-Fixing, and Corporate Leniency Policy: What Doesn't Kill Us Makes Us Stronger," University of Oregon, 2001.

[12] Friederiszick, Hans W. and Frank P. Maier-Rigaud, “Triggering Inspections Ex Officio: Moving Beyond a Passive EU Cartel Policy,” *Journal of Competition Law and Economics*, 4 (2008), 89-113.

[13] Ghosal, Vivek and Joseph Gallo, “The Cyclical Behavior of the Department of Justice’s Antitrust Enforcement Activity,” *International Journal of Industrial Organization,* 19 (2001), 27-54.

[14] Harrington, Joseph E., Jr., “Behavioral Screening and the Detection of Cartels,” in *European Competition Law Annual 2006: Enforcement of Prohibition of Cartels*, Claus-Dieter Ehlermann and Isabela Atanasiu, eds., Oxford: Hart Publishing, 2007.

[15] Harrington, Joseph E., Jr., “Optimal Corporate Leniency Programs,” *Journal of Industrial Economics*, 56 (2008), 215-246.

[16] Harrington, Joseph E., Jr., “When is an Antitrust Authority not Aggressive Enough in Fighting Cartels?,” *International Journal of Economic Theory*, 7 (2011a), 39-50.

[17] Harrington, Joseph E., Jr., “Corporate Leniency Programs when Firms have Private Information: The Push of Prosecution and the Pull of Pre-emption,” Johns Hopkins University, 2011b (*Journal of Industrial Economics*, forthcoming).

[18] Harrington, Joseph E., Jr., “Corporate Leniency with Private Information: An Exploratory Example,” in *Recent Advances in the Analysis of Competition Policy and Regulation*, J. Harrington and Y. Katsoulakos, eds., Cheltenham, UK: Edward Elgar, 2012.

[19] Harrington, Joseph E., Jr. and Myong-Hun Chang, “Modelling the Birth and Death of Cartels with an Application to Evaluating Antitrust Policy,” *Journal of the European Economic Association*, 7 (2009), 1400-1435.

[20] Hinloopen, Jeroen and Adriaan R. Soetevent, “Laboratory Evidence on the Effectiveness of Corporate Leniency Programs,” *RAND Journal of Economics*, 39 (2008), 607-616.

[21] Hyytinen, Ari, Frode Steen, and Otto Toivanen, “Cartels Uncovered,” Centre for Policy Economic Research, Discussion Paper No. 7761, March 2010.

[22] Jeffrey, R. L., "The Continuity of a Function Defined by a Definite Integral," *The American Mathematical Monthly*, 32 (1925), 297-299.

[23] Klein, Gordon, “Cartel Destabilization and Leniency Programs - Empirical Evidence,” ZEW Discussion Paper No. 10-107, 2010.

[24] Kwoka, John, “Commitment to Competition: An Assessment of Antitrust Agency Budgets since 1970,” *Review of Industrial Organization,* 14 (1999), 295-302.

[25] Miranda, Mario J. and Paul L. Fackler, *Applied Computational Economics and Finance*, Cambridge, Mass.: The MIT Press, 2002.

[26] Miller, Nathan, "Strategic Leniency and Cartel Enforcement," *American Economic Review*, 99 (2009), 750-768.

[27] Motta, Massimo and Michele Polo, "Leniency Programs and Cartel Prosecution," *International Journal of Industrial Organization*, 21 (2003), 347-379.

[28] Riley, Alan, "Developing Criminal Cartel Law: Dealing with the Growing Pains," *Competition Law Review*, 4 (2007), 1-6.

[29] Rotemberg, Julio J. and Garth Saloner, "A Supergame-Theoretic Model of Price Wars During Booms," *American Economic Review*, 76 (1986), 390-407.

[30] Spagnolo, Giancarlo, "*Divide et Impera*: Optimal Deterrence Mechanisms Against Cartels and Organized Crime," University of Mannheim, 2003.

[31] Spagnolo, Giancarlo, "Leniency and Whistleblowers in Antitrust," in *Handbook of Antitrust Economics*, Paolo Buccirossi, ed., Cambridge, Mass.: The MIT Press, 2008.

[32] Veljanovski, Cento, "European Commission Cartel Prosecutions and Fines, 1998-2007," Case Associates, 2007.

[33] Zhou, Jun, "Evaluating Leniency and Modeling Cartel Durations: Time-Varying Policy Impacts and Sample Selection," Bonn University, September 2011.