When Can We Expect a Corporate Leniency Program to Result in Fewer Cartels?*

Journal of Law and Economics, forthcoming

Joseph E. Harrington, Jr.
Dept. of Bus. Econ. & Public Policy
The Wharton School
University of Pennsylvania
Philadelphia, PA 19104
harrij@wharton.upenn.edu

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

13 January 2015

Abstract

Leniency programs have become widespread and are generally quite active as reflected in the number of applications. What is not well-understood is how they affect the number of cartels. This paper develops and explores a theoretical framework to help understand when leniency programs are likely to be effective in reducing the presence of cartels. Plausible conditions are derived whereby a leniency program can result in more cartels. On a more positive note, we identify situations and policies that a competition authority can pursue that will make it more likely that a leniency program will have the intended effect of reducing the number of cartels.

*For their valuable comments on a previous version of this paper, we thank an anonymous referee, 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, U. of Texas at Dallas, U.of North Carolina at Chapel Hill, Northwestern University-KGSM, Universitat Pompeu Fabra, CEMFI, Universität Düsseldorf (DICE), VU Amsterdam & The Tinbergen Institute, and Harvard Law School. The first author gratefully acknowledges the financial support of the National Science Foundation (SES-0516943, SES-1148129) and Santander through a Cátedras de Excelencia. He also expresses his gratitude to the faculty of the Departamento de Economía of the Universidad Carlos III de Madrid for providing a most collegial and stimulating environment.

1 Introduction

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 report the cartel and fully cooperate. Since its revision in 1993, the Program has been flush with applications. Deputy Assistant Attorney General Scott Hammond noted in 2005 that "the revised Corporate Leniency Program has resulted in a surge in amnesty applications. Under the new policy, the application rate has jumped to roughly two per month" (see Hammond [2005], p. 10). Furthermore, he commented: "The extraordinary success of the Division's leniency program has generated widespread interest around the world" (see Hammond [2005], p. 10). That is indeed the case as the steady flow of leniency applications in the U.S. led the European Commission (EC) to institute its own leniency program in 1996 and a decade later 24 out of 27 EU members had one. Globally, leniency programs are now present in more than 50 countries and jurisdictions.[1]

In many of these countries, leniency programs have similarly been active and viewed as successful. In South Africa, which put in place its program in 2004, the number of applicants were flowing in at a rate of about three per month by 2009-10, which even exceeded that in the U.S. A week prior to Spain's institution of its leniency program in June 2008, cartelists were literally lining up outside the doors of the Comisión Nacional de la Competencia's offices in order to be the first from their cartel to apply for leniency. In Germany, the Cartel Office noted: "The first version of the Leniency Programme was already a success. This can be seen by the number of leniency applications filed: Between 2000 and 2005 a total of 122 leniency applications were filed" (see Bundeskartellamt [2010], p.19).

It is clear that many leniency programs have sparked numerous applications. It is also clear that one can identify specific cases for which a leniency program was responsible for the discovery of the cartel and was instrumental in the successful prosecution of the cartel. What is far less clear, however, is whether leniency programs have been successful in the sense that these economies are populated by fewer cartels. Ultimately, success is to be measured by a small number of cartels, not a large number of leniency applications.

In light of the widespread adoption and usage of leniency programs, there is a vast and growing body of scholarly work intended to examine the effect of these programs (see Spagnolo [2008] for a review of some of the early work). Starting with the pioneering paper of Motta and Polo (2003), theoretical analyses include Ellis and Wilson (2001), Spagnolo (2003), Motchenkova (2004), Aubert, Kovacic, and Rey (2006), Chen and Harrington (2007), Harrington (2008), Har-

[1] For a list of countries with leniency programs, see Borrell, Jiménez, and García (2014) who also estimate how leniency programs have changed the perceptions of managers.

rington and Chang (2009), Houba, Motchenkova, and Wen (2009), Silbye (2010), Choi and Gerlach (2012), Lefouili and Roux (2012), Sauvagnat (2012), Bos and Wandschneider (2013), Chen and Rey (2013), Gärtner (2013), Marshall, Marx, and Mezzetti (2013), Blatter, Winand Emons, and Sticher (2014), and Marx and Mezzetti (2014). The general conclusion of this body of work is that leniency programs make collusion more difficult.

A common feature to all of these models is the assumption that the introduction of a leniency program does not impact enforcement through non-leniency means. Non-leniency enforcement is modelled as the probability that a cartel is discovered, prosecuted, and convicted in the absence of a member having entered the leniency program. As a cartel member will apply for leniency only if it believes that doing so is better than running the risk of being caught and convicted, non-leniency enforcement is integral to inducing firms to apply for leniency. If this probability is low then few cartel members will use the leniency program, while if the probability is high then cartel members will race to apply for leniency. The impact of a leniency program is then intrinsically tied to the level of non-leniency enforcement.

Next note that it is natural to expect that the introduction of a leniency program will affect the level of non-leniency enforcement. A leniency program may cause a competition authority's scarce resources and attention to shift from non-leniency cases to leniency cases. However, this doesn't necessarily imply that non-leniency enforcement is weaker. If a leniency program is successful in resulting in fewer cartels, there will be fewer non-leniency cartel cases, in which case the authority may still have ample resources to effectively prosecute them. Furthermore, a competition authority can adjust its enforcement policy in response to what is occurring with leniency applications. Thus, while we expect non-leniency enforcement to change when a leniency program is put in place, it isn't clear whether it'll be weakened or strengthened.

The objective of the current paper is to develop and explore a theoretical framework to understand when leniency programs are likely to be effective in reducing the presence of cartels. Its primary innovation is in providing a more comprehensive assessment of how a leniency program affects the activity and efficacy of a competition authority by taking account of its impact on the entire portfolio of cases; both those generated through leniency applications and through more traditional methods. Our model is the first to examine the effect of a leniency program while endogenizing non-leniency enforcement.

Contrary to existing results in the literature and the general impression of practitioners, we find that a leniency program can result in *more* cartels, and this can occur at the same time that a leniency program is generating many applications. On a more positive note, we also identify situations and policies that a competition authority can pursue that will make it more likely that a leniency program will have the intended effect of reducing the number of cartels.

Before moving on, it is useful to note that, in addition to a burgeoning theoretical literature on leniency programs, there is a growing body of experimental work. Research here includes Apesteguia, Dufwenberg, and Selten (2007), Hinloopen and Soetevent (2008), Hamaguchi, Kawagoe, and Shibata (2009), Dijkstra, Haan, and Schoonbeek (2011), and Bigoni, Fridolfsson, Le Coq, and Spagnolo (2012). These experimental studies generally find 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 is an increasing number of empirical studies that measure the impact of leniency programs but are decidedly mixed and tentative in their findings. Miller (2009) examines the impact of the leniency program in the U.S., Choi (2009) and Koh and Jeong (2013) consider the program in Korea; Stephan (2008), Brenner (2009), Klein (2010), and Zhou (2011) investigate the impact of the European Commission's leniency program; and Dong, Massa, and Žaldokas (2014) engage in a cross-country analysis.

In the next section, the model is presented. In Section 3, the conditions determining the equilibrium cartel rate are derived. The impact of a leniency program on the cartel rate when non-leniency enforcement is fixed is examined in Section 4. While those results are of intrinsic interest as a benchmark, they are primarily an intermediate step towards endogenizing non-leniency enforcement. Section 5 delivers the main contribution of the paper which is a characterization of how a leniency program impacts the cartel rate when non-leniency enforcement is allowed to adjust to a competition authority experiencing a flow of leniency applications. Section 6 concludes. Proofs are in the online appendix.

2 Model

The modelling strategy is to construct a birth-and-death process for cartels in order to generate an average cartel rate for a population of industries, and then to assess how the introduction of a leniency program influences that cartel rate. We build on the birth-and-death process developed in Harrington and Chang (2009) by introducing a leniency program, endogenizing the intensity of non-leniency enforcement, and allowing a competition authority (CA) to decide on its caseload. In this manner, the full effect of a leniency program can be assessed.[2]

[2] While a leniency program is briefly considered in Harrington and Chang (2009), that model assumes - like the remainder of the literature - that non-leniency enforcement is fixed. As hopefully this paper will convince the reader, it is a technically and economically substantive extension of the Harrington-Chang model to endogenize non-leniency enforcement.

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 π (see Rotemberg and Saloner [1986] the same informational setting). 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 $H : [\underline{\pi}, \overline{\pi}] \to [0, 1]$ 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 Table 1.

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 $(\alpha, \eta) = \left(\frac{4n}{(n+1)^2}, \frac{(n+1)^2}{4n}\right)$.[3]

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 have the *opportunity to collude* but do so if and only if (iff) 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 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 CA may pursue an investigation; this can only occur if, in the current period, the cartel was either active or shut down and no firm applied for leniency. 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

[3] 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.

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 equal to the length of time for which it has been active; such an extension 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.[4] As the CA will not be presumed to manipulate F, one can suppose that penalties are already set at their maximum level.

In addition to being discovered by the CA, a cartel can be uncovered because one of its members comes forth under the corporate leniency program. 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 a firm, 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).

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) 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

[4] A more standard assumption in the literature is to assume F is fixed which is certainly simpler but less realistic than our specification. All of our qualitative results hold when F is fixed.

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. The timing of events is summarized in Figure 1.

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 simplifies the analysis.

2.2 Enforcement Technology

Non-leniency enforcement refers to σ which is the probability that a cartel pays penalties without one of its members having entered the leniency program. σ is the compound probability that: 1) a cartel is discovered by the CA; 2) the CA decides to investigate a discovered cartel; and 3) the CA is successful in its investigation and penalties are levied. The probability of discovery $q \in (0, 1)$ is presumed to be exogenous and to come from customers, uninvolved employees, the accidental discovery of evidence through a proposed merger, and so forth. Of those reported cases, the CA controls the fraction to investigate which is denoted $r \in (0, 1]$.[5] Finally, of those cases discovered and prosecuted, the CA is successful in a fraction $s \in [0, 1]$ of them where, as described next, s is determined by the CA's caseload.

Stacking the model in favor of a leniency program, it is assumed that leniency cases are won for sure, while a non-leniency case is won with probability s. What is important is that the likelihood of a conviction in a leniency case exceeds that in a non-leniency case which is compelling given that the former has a cartel member acting as a witness. In prosecuting non-leniency cases, the CA is faced with a resource constraint: the larger is its caseload, the fewer resources are applied to each case and the lower is the probability of winning any individual case. More specifically, it is assumed $s = p(\lambda L + R)$ where L is the mass of leniency cases and R is the mass of non-leniency cases handled by the CA (and $R = q \times r \times C$ where C is the mass of cartels). It is assumed $\lambda \in (0, 1]$ because leniency cases

[5] There is no loss of generality in assuming $r > 0$ because if $r = 0$ then there is effectively no enforcement given there are no non-leniency cases and, in addition, firms will not apply for leniency as there is no prospect of being convicted.

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. $p:[0,1]\rightarrow[0,1]$ 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 $\sigma=q\times r\times s=q\times r\times p\left(\lambda L+R\right)$.

Key to the analysis is that the CA implicitly faces a resource constraint. While, in practice, a CA can move around resources to handle additional cartel activity by, for example, shifting lawyers and economists from merger cases to cartel cases, there is a rising opportunity cost in doing so and that ought to imply that resources per cartel case will decline with the number of cartel cases. Of course, CA officials can lobby their superiors 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. To the contrary, the case activity of the Antitrust Division of the U.S. Department of Justice is countercyclical (Ghosal and Gallo, 2001) but its budget is pro-cyclical for it is increasing in GDP (Kwoka, 1999).

The final element of the enforcement technology to discuss is the determination of how many cases the CA takes on. If the CA maximizes welfare (which, in this model, means minimizing the cartel rate) and can freely choose how many leniency and non-leniency cases to take on then a leniency program cannot raise the cartel rate because the CA has the option of replicating its policy in the absence of a leniency program. However, our objective is not to examine such an idealized setting but rather to determine whether a leniency program can raise the cartel rate assuming a plausible description of CA practices; and, if a leniency program can raise the cartel rate, to identify implementable policies that would make it more likely that a leniency program lowers the cartel rate.

With that perspective, the CA is assumed to accept all leniency applications and thereby prosecute all leniency cases. In the context of our model, this assumption is consistent with practice. According to the head of the cartels directorate unit of the European Commission: "As a matter of practice we pursue virtually all cases where we think we can make the case. (I'm tempted to say 'pursue all cases' but a good lawyer rarely talks in absolutes.)"[6] Given leniency cases are won for sure, the CA then prosecutes all leniency cases in our model. A second motivation for this assumption is that it is likely to be an implication of a plausible specification of CA preferences. However a CA is rewarded, those rewards are presumably based on observable measures of performance such as the number of cases won, the percentage of cases won, and the amount of fines collected.[7] As le-

[6] Email from Kevin Coates to Joseph Harrington, July 29, 2014.

[7] Consistent with a focus on observable performance measures, Assistant Attorney General Thomas Barnett of the Antitrust Division of the U.S. Department of Justice stated in his opening remarks during Congressional testimony (September 25, 2007) that: "The Division set a record for the most jail time imposed (almost 30,000 jail days) [and] obtained the second highest amount of fines in the Division's history (over $630 million)."

niency cases are sure to add to a CA's performance by delivering more convictions and contributing to a high prosecutorial success rate, it is reasonable to suppose that they pursue all (winnable) leniency cases. While we believe it is reasonable to assume that the CA takes on all leniency cases, our results are robust as long as enough leniency cases are taken on, which is shown at the end of Section 5.

The more problematic decision for the CA is how many non-leniency cases to prosecute given that they are more difficult to win. Rather than take a position on the CA's preferences (which we leave to future research), results are derived for all values of r. For example, sufficient conditions will be derived for a leniency program to raise the cartel rate regardless of the fraction of non-leniency cases that a CA takes on. Thus, results do not depend on the CA's preferences and apply whether the CA chooses its non-leniency caseload to maximize the number of convictions or minimize the cartel rate or achieve some other objective.[8]

3 Equilibrium Conditions

In this section, we describe the conditions determining the equilibrium frequency with which industries are cartelized. Given that there are several steps in the construction of equilibrium conditions, it may prove beneficial to the reader to provide an 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

[8] In discussing the behavior of the CA, it is worth noting that Motta and Polo (2003) assume enforcement expenditure is set optimally when 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.

on 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:

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

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 σ. Given σ^*, the equilibrium cartel rate is $C(\sigma^*)$.

By way of comparison, the model in Harrington and Chang (2009) involved the two nested fixed point problems in steps 1 and 2 when σ is fixed. The current model embeds that problem in a third fixed point problem in order to endogenize σ (step 3). We believe this extension is both technically and economically substantive in that it introduces fundamentally new forces relevant to assessing the effect of an antitrust policy.

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:

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

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); that is, when $\theta\gamma(Y-\alpha\mu) < \sigma\gamma(Y-\alpha\mu)$ 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.[9]

[9] 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 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 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:

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

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 and the expected penalty is $\omega\gamma(Y-\alpha\mu)$.

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 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.

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

$$\phi^*(\sigma,\eta) \equiv \max\left\{\min\left\{\phi\left(Y^*(\sigma,\eta),\sigma,\eta\right),\overline{\pi}\right\},\underline{\pi}\right\}, \tag{5}$$

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).

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 $(1-\sigma)H(\phi^*(\sigma,\eta))$.

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

$$\begin{aligned} NC(\sigma,\eta) &= NC(\sigma,\eta)\left[(1-\kappa)+\kappa(1-H(\phi^*))+\kappa\sigma H(\phi^*)\right] \\ &\quad +\left[1-NC(\sigma,\eta)\right]\left[(1-H(\phi^*))+\sigma H(\phi^*)\right] \end{aligned} \tag{6}$$

Examining the RHS of (6), 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 $\kappa\left(1-H\left(\phi^*\right)\right)$ will have the opportunity but, due to a high realization of π, find it is not incentive compatible to collude, while a fraction $\kappa\sigma H\left(\phi^*\right)$ will cartelize and collude but then are discovered by the CA. Of the industries that were colluding in the previous period, which have mass $1-NC\left(\sigma,\eta\right)$, a fraction $1-H\left(\phi^*\right)$ will collapse for internal reasons and a fraction $\sigma H\left(\phi^*\right)$ will instead be shut down by the authorities.

Solving (6) for $NC\left(\sigma,\eta\right)$:

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

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

$$C\left(\sigma,\eta\right)\equiv 1-NC\left(\sigma,\eta\right)=\frac{\kappa\left(1-\sigma\right)H\left(\phi^*\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right)H\left(\phi^*\left(\sigma,\eta\right)\right)}. \tag{8}$$

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 $C\left(\sigma\right)$, is then defined by:

$$C\left(\sigma\right)\equiv\int_{\underline{\eta}}^{\overline{\eta}}C\left(\sigma,\eta\right)g\left(\eta\right)d\eta=\int_{\underline{\eta}}^{\overline{\eta}}\left[\frac{\kappa\left(1-\sigma\right)H\left(\phi^*\left(\sigma,\eta\right)\right)}{1-\left(1-\kappa\right)\left(1-\sigma\right)H\left(\phi^*\left(\sigma,\eta\right)\right)}\right]g\left(\eta\right)d\eta. \tag{9}$$

3.3 Equilibrium Non-Leniency Enforcement

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:

$$L\left(\sigma\right)=\begin{cases}0 & \text{if } \sigma\leq\theta\\ \int_{\underline{\eta}}^{\overline{\eta}}\left(1-H\left(\phi^*\left(\sigma,\eta\right)\right)\right)C\left(\sigma,\eta\right)g\left(\eta\right)d\eta & \text{if } \theta<\sigma\end{cases} \tag{10}$$

In (10), 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. That it is dying cartels that apply for leniency is consistent with the European Commission experience.[10, 11]

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

$$R(\sigma) = qr\left[C(\sigma) - L(\sigma)\right] = \begin{cases} qrC(\sigma) & \text{if } \sigma \leq \theta \\ qr\int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma,\eta))\, C(\sigma,\eta)\, g(\eta)\, d\eta & \text{if } \theta < \sigma \end{cases} \tag{11}$$

If the leniency program is never used (which is when $\sigma \leq \theta$), then the mass of cases being handled by the CA is $qrC(\sigma)$. 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 $\int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma,\eta))\, C(\sigma,\eta)\, g(\eta)\, d\eta$.

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:

$$\sigma = \Psi(\sigma) \equiv \begin{cases} qrp(qrC(\sigma)) & \text{if } \sigma \leq \theta \\ qrp\Big(\lambda \int_{\underline{\eta}}^{\overline{\eta}} (1 - H(\phi^*(\sigma,\eta)))\, C(\sigma,\eta)\, g(\eta)\, d\eta & \text{if } \theta < \sigma \\ \quad + qr\int_{\underline{\eta}}^{\overline{\eta}} H(\phi^*(\sigma,\eta))\, C(\sigma,\eta)\, g(\eta)\, d\eta\Big) & \end{cases} \tag{12}$$

where we have substituted for L using (10) and R using (11).[12] If there are multiple solutions to (12) then it is assumed the maximal one is selected.[13]

4 Impact of a Leniency Program When Non-Leniency Enforcement is Exogenous

In this section, results are derived under the standard assumption in the literature that non-leniency enforcement is fixed. These results are a necessary intermediate

[10] EC official Olivier Guersent expressed a concern that leniency applications were coming from dying cartels at the 11th Annual EU Competition Law and Policy Workshop: Enforcement of Prohibition of Cartels in Florence, Italy in June 2006. Jun Zhou conveyed in a conversation that only 13 out of 110 EC cases with a leniency awardee (over 1996-2012) involved applications before the death of the cartel.

[11] 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 Harrington (2012, 2013) and Marx and Mezzetti (2014).

[12] 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.

[13] 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^*.

step towards deriving the paper's main results for when non-leniency enforcement is endogenized, but also serve as a benchmark for highlighting how an evaluation of a leniency program significantly changes when a more comprehensive analysis is performed.

To begin, consider the cartel rate function $C(\sigma)$; that is, the cartel rate that results for a given level of non-leniency enforcement σ. Theorem 1 is a re-statement of a result in Harrington and Chang (2009) and shows that when firms assign a higher probability to the CA discovering, prosecuting, and convicting cartels then a smaller fraction of industries are cartelized. This result is derived for when the penalty multiple γ is not too high and, in the ensuing analysis, it is assumed (without being stated) that γ is such that Theorem 1 applies.[14,15]

Theorem 1: $\exists \widehat{\gamma} > 0$ 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.

Next let us consider how a leniency program affects the cartel rate function; that is, whether a leniency program results in a higher or lower cartel rate for a given value of σ. For this purpose, all functions will be subscripted with θ when there is a leniency program with parameter θ; for example, $C_\theta(\sigma)$ denotes the cartel rate function in that case.[16] Similarly, all functions will be subscripted with NL when there is no leniency program; for example, $C_{NL}(\sigma)$ denotes the cartel rate function.

Theorem 2 shows that a leniency program does not raise the cartel rate function, and it reduces the cartel rate function if, in the absence of a leniency program, there is a positive measure of industries that cannot fully collude and a positive measure that can collude (Assumption A1).[17] In sum, for a given level of non-leniency enforcement, a leniency program results in fewer cartels.

Assumption A1: 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 $\phi^*_{NL}(\sigma, \eta) > \underline{\pi}$.

[14] It is also the case that γ must not be too high if there is to be some collusion in equilibrium. However, we have not ruled out the possibility that γ could be low enough for $C(\sigma) > 0$ but is not low enough for $C'(\sigma) < 0$ (that is, Theorem 1 applies) though we have no reason to think that to be true.

[15] The analogous result in Harrington and Chang (2009) assumes σ is sufficiently small which we do not want to do here because σ will later be endogenized. Instead, results are derived for when the penalty multiple γ is sufficiently small. This involves a straightforward modification of the proof in Harrington and Chang (2009) and thus is not included here though is available on request.

[16] 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. To reduce notational clutter, we will suppress ω.

[17] A1 ensures that the cartel rate is positive but not maximal, and rules out extreme 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.

Theorem 2: If $\sigma \in (\theta, \omega)$ then $C_{NL}(\sigma) \geq C_\theta(\sigma)$ and if A1 holds then $C_{NL}(\sigma) > C_\theta(\sigma)$.

Prior to explaining why Theorem 2 is true, let us interpret and motivate the restriction $\sigma \in (\theta, \omega)$. If $\sigma > \theta$ then a firm that contemplates deviating from a cartel would apply for leniency because doing so reduces the 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 the unique equilibrium play.[18] In that situation, 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 when it deviates or when the cartel collapses and, in the latter situation, expected penalties are higher compared to when there is no leniency program. If instead $\sigma < \theta$ then firms would never apply for leniency in which case a leniency program is ineffectual. If $\sigma > \omega$ then, upon collapse of the cartel, firms would use leniency but, relative to the absence of a leniency program, expected penalties are *lower*. In that situation, it is not difficult to derive conditions such that a leniency program raises the cartel rate. Thus, $\sigma \in (\theta, \omega)$ is the relevant domain for our analysis in that a leniency program is active and it is properly designed so that it raises expected penalties. Given that σ will be endogenized, this leaves open the question of whether its equilibrium value lies in (θ, ω). Note that $\sigma\,(> 0)$ is sure to exceed θ when $\theta = 0$ (so there is full leniency), and that case is the focus of Section 5. If all firms applying for leniency gives each an equal chance of receiving it, then $\omega = \frac{n-1+\theta}{n} \geq \frac{n-1}{n} \geq \frac{1}{2}$. Thus, if $q < 1/2$ then $\omega \geq 1/2 > \sigma\,(= qrs)$. Hence, if $\theta = 0$ and $q < 1/2$ then $\sigma \in (\theta, \omega)$.[19]

Turning to the interpretation of Theorem 2, it shows that a leniency program reduces the frequency of cartels when non-leniency enforcement is fixed. Let us summarize the forces that are the basis for that result (also see these forces in Motta and Polo [2003], Spagnolo [2003], and Harrington [2008]). A leniency program increases the payoff to cheating because now a firm can reduce its penalty by simultaneously applying for leniency. This shrinks the set of market conditions for which collusion is stable and thereby reduces expected cartel duration and the value to colluding. A leniency program also reduces the value to colluding because, upon collapse, firms race for leniency and that results in higher expected

[18] If all other firms were expected not to apply then a firm's penalty from applying is $\theta\gamma(Y - \alpha\mu)$ and from not applying is $\sigma\gamma(Y - \alpha\mu)$. If $\sigma > \theta$ then a firm prefers to apply in which case all firms not applying is not an equilibrium. Given that it is optimal to apply when one or more other firms apply, equilibrium play must then involve all firms applying when $\sigma > \theta$.

[19] Let us also note that $\sigma \in (\theta, \omega)$ is required only for some results and only because it is a sufficient condition for the cartel rate function to be higher without a leniency program: $C_{NL}(\sigma) > C_\theta(\sigma)$; see Theorem 2. When $\sigma \in (\theta, \omega)$ is needed, that condition is stated as part of the theorem. Of particular note, it is is not required for Theorem 4 which provides sufficient conditions for a leniency program to raise the cartel rate.

penalties. Due to the lower value to colluding, either a cartel no longer forms or it has shorter duration and this translates into a lower aggregate cartel rate.

5 Impact of a Leniency Program When Non-Leniency Enforcement is Endogenous

We are now prepared to provide a comprehensive assessment of how a corporate leniency program influences the cartel rate. Our plan is to address the following questions: Taking account of the effect of a leniency program on a CA's prosecutions of both leniency and non-leniency cases, can a leniency program raise the cartel rate? If it can, under what circumstances can we be assured that a leniency program is lowering the cartel rate?

The analysis will focus on when the leniency program provides full leniency to the first firm to come forward ($\theta = 0$). This is a natural case to consider because almost all leniency programs waive all government penalties to the first firm to come forward prior to the start of an investigation.[20] We also focus on this case for technical reasons related to the existence of equilibrium.

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 equilibrium values for non-leniency enforcement are σ^*_{NL} and σ^*_L in which case the equilibrium cartel rates are $C_{NL}(\sigma^*_{NL})$ and $C_L(\sigma^*_L)$.

With this notation, we can summarize the task before us. Theorem 2 showed that a leniency program lowers the cartel rate given a value for σ: $C_L(\sigma) < C_{NL}(\sigma)$. Whether a leniency program raises or lowers the cartel rate 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 because, by Theorem 1, cartel rate functions are decreasing: $C_L(\sigma^*_L) < C_{NL}(\sigma^*_{NL})$. If instead a leniency program weakens non-leniency enforcement - $\sigma^*_L < \sigma^*_{NL}$ - then the ultimate impact on the cartel rate depends on the extent to which a leniency program reduces non-leniency enforcement.

In Section 5.1, it is shown that an equilibrium cartel rate exists for when there is no leniency program or a full leniency program. In Section 5.2, it is shown that a leniency program can raise the cartel rate. In Section 5.3, conditions are derived for a leniency program to lower the cartel rate. While the analysis largely focuses on how a leniency program affects the frequency of cartels, Section 5.4 shows the

[20] In the U.S., a firm that receives amnesty is still liable for single customer damages so leniency is not full. Most other jurisdictions do not have customer damages in which case government fines encompass the entirety of penalties and, therefore, leniency is approximately full.

differential impact of a leniency program across industries; specifically, a leniency program can make collusion more difficult in some industries but less difficult in other industries.

5.1 Existence of an Equilibrium Cartel Rate

The equilibrium level of non-leniency enforcement σ^* is a fixed point to Ψ which is defined in (12). σ^* has the property that if firms believe 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 probability is indeed σ^*.

Theorem 3: For $\theta \in \{0,1\}$, $\exists \widehat{\gamma} > 0$ such that if $\gamma \in [0, \widehat{\gamma})$ then σ^* exists.

While we will only be assessing the impact of providing full leniency, we do not believe the intuition behind the results is tied to leniency being full. We conjecture that, as long as σ^* exists, results will go through if, in equilibrium, firms use leniency in response to cartel collapse (that is, $\theta < \sigma^*$) and penalties are higher as a result of leniency (that is, $\sigma^* < \omega$).

5.2 A Leniency Program Increases the Cartel Rate

In this section, we show that a leniency program can be counter-productive. When penalties are not severe enough and the amount of resources saved by prosecuting a leniency case are not large enough then the introduction of a leniency program *raises* the cartel rate.

Theorem 4: Assume

$$\int_{\underline{\eta}}^{\overline{\eta}} \left(1 - H\left(\phi_{NL}^*\left(\sigma_{NL}^*, \eta\right)\right)\right) C_{NL}\left(\sigma_{NL}^*, \eta\right) g\left(\eta\right) d\eta > 0 \tag{13}$$

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

In understanding the forces that drive this result, first note that a leniency program can affect the cartel rate by *disabling* active cartels (that is, shutting them down) and by *deterring* new cartels from forming. A leniency program can have a perverse effect because, while it generally promotes deterrence, it can actually result in fewer cartels being shut down.

Prior to the introduction of a leniency program, the CA is discovering, prosecuting, and convicting cartels through non-leniency means. While some of the cartels that are convicted will just so happen to have internally collapsed, many of them will have been active in which case it is their prosecution and conviction that shuts the cartel down. When a leniency program is introduced, cartels that collapse race for leniency and these leniency applications comprise part of the caseload of the CA. Of particular note is that leniency cases are coming from *dying cartels* and thus prosecution of them is not shutting down an active cartel. However, these leniency cases add to the CA's caseload and thereby result in less success in prosecuting non-leniency cases which, if it had led to a conviction, would have disabled a well-functioning cartel. In essence, leniency cases - which do not shut down an active cartel - are crowding out non-leniency cases - which generally do shut down active cartels. If leniency cases do not save much in terms of prosecutorial resources (that is, λ is not sufficiently below one) then this crowding out effect is significant and the end result is that many fewer cartels are shut down when there is a leniency program.

This is not the end of the story, however. Due to the leniency program, a dying cartel is now assured of paying penalties because one of its members will enter the leniency program and aid the CA in obtaining a conviction. In contrast, without a leniency program, only a fraction of those cartels would have been discovered and penalized (specifically, the fraction is σ^*_{NL}). Thus, a leniency program is raising the expected penalties for a cartel in the event of its death which serves to deter some cartels from forming. However, if penalties are not large enough (that is, γ is not sufficiently great) then the additional cartels deterred due to the leniency program is small in comparison to the reduction in the number of cartels shut down because leniency cases crowd out non-leniency cases. As a result, on net, the cartel rate is higher. Thus, 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.

Theorem 4 shows that, for any value of r (which recall is the fraction of possible non-leniency cases that the CA chooses to prosecute), the cartel rate is higher with a leniency program:[21]

$$C_L\left(qrs^*_L\left(r\right)\right) > C_{NL}\left(qrs^*_{NL}\left(r\right)\right),\ \forall r > 0. \tag{14}$$

Recall that $\sigma = qrs$, and in (14) we have made explicit the dependence of the conviction rate s on r. If the CA chooses its caseload in order to minimize the cartel rate, the optimal prosecution policy with and without a leniency program, respectively, is

$$r^*_L \in \arg\min_{r\in[0,1]} C_L\left(qrs^*_L\left(r\right)\right),\ r^*_{NL} \in \arg\min_{r\in[0,1]} C_{NL}\left(qrs^*_{NL}\left(r\right)\right).$$

[21] Theorem 4 does require that for any value of r, (13) is satisfied which is not very restrictive because whether a cartel collapses is partly due to forces unrelated to the CA. For example, even in the absence of a CA, all cartels will collapse with positive probability when $\overline{\pi}$ is sufficiently high and $\underline{\eta} > 1$.

It then follows from (14),

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

This leads to the following corollary.

Corollary 5: Under the conditions of Theorem 4, a leniency program raises the cartel rate even if the competition authority is choosing its caseload to minimize the cartel rate.

In concluding this section, let us argue that the sufficient conditions in Theorem 4 for a leniency program to raise the cartel rate could plausibly hold in some jurisdictions. A leniency program raises the cartel rate when: 1) a leniency case takes up a reasonable amount of resources so that there is a crowding out of resources for non-leniency cases; and 2) penalties are not so severe that they significantly deter cartel formation.

With regards to the first condition, the Directorate General Competition (DG Comp) of the European Commission was initially overwhelmed with leniency applications which could well have significantly limited the availability of resources for prosecuting other cases (see Riley [2007], pp. 1-2.):

> 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 impact of a leniency program on enforcement through other means was a concern emphasized in Friederiszick and Maier-Rigaud (2008). Both authors were members of DG Comp and their paper recommended that the DG Comp step up non-leniency enforcement methods such as being active in detecting cartels. Consistent with these views, Kai-Uwe Kühn, who was Chief Economist of the European Commission, expressed at the Searle Research Symposium in September 2010 that leniency cases seem to be as long and involved as non-leniency cases. It is then quite plausible that jurisdictions with an active leniency program could experience weakened non-leniency enforcement because of the crowding out of non-leniency methods of enforcement.

Turning to the second condition, many countries have set legal caps on fines that, for cartels in reasonably-sized markets, are likely to be far below the incremental profit from colluding and thus do little to deter cartel formation. For example, the maximum penalty in Chile is around $25 million per defendant. By way of comparison, there is currently a case against a suspected cartel in the wholesale chicken market which has annual sales on the order of $1 billion. Even more paltry are caps of around \$7 million in Mexico (at least until 2011 when it was increased) and \$5 million in Japan (see Report to the 7th ICN Annual Conference [2008]).

In sum, there is at least anecdotal evidence to suggest that leniency cases could be reducing the rate at which a CA shuts down active cartels because resources are taken away from non-leniency enforcement, and that penalties could be at sufficiently low levels such that the threat of a race for leniency does not substantively deter cartel formation. It is then plausible that some jurisdictions could be experiencing a higher cartel rate because of having introduced a leniency program.

5.3 A Leniency Program Decreases the Cartel Rate

As just shown, a leniency program is not assured of reducing the frequency of cartels; it can raise the cartel rate. One tactic that a CA can take to avoid this outcome and ensure that a leniency program serves the cause of fighting cartels is to set up a procedure that will expeditiously handle leniency cases. Theorem 6 establishes that if leniency cases save sufficient resources - relative to non-leniency cases - then a leniency program will lower the cartel rate.

Theorem 6: If $\sigma^*_{NL} \in (0, \omega)$ then there exists $\widehat{\lambda} > 0$ such that if $\lambda \in \left[0, \widehat{\lambda}\right]$ then the cartel rate is weakly lower with a leniency program and if A1 holds then the cartel rate is strictly lower with a leniency program.

While this result is not particularly surprising, it is important to understand why it is true. It is not just that the crowding out of non-leniency enforcement is reduced - and thus non-leniency enforcement does not fall as much - but rather that non-leniency enforcement can actually be higher.

Key to a leniency program raising the cartel rate is that leniency cases are consuming valuable CA resources which detracts from non-leniency enforcement. However, if leniency cases can be handled with few resources then they will not crowd out many non-leniency cases. Though there will still be some crowding out - which would suggest non-leniency enforcement would still be harmed and thus fewer cartels are shut down by the CA - there is a mitigating effect coming from the enhanced deterrence of cartel formation due to a leniency program. By raising the penalties that a cartel can expect to pay when it collapses (and cartel members subsequently race for leniency), a leniency program results in fewer cartels forming. With fewer cartels forming, there will be fewer non-leniency cases. Thus, non-leniency enforcement could be stronger (that is, $\sigma^*_L > \sigma^*_{NL}$) because, while there are fewer resources for non-leniency cases due to the presence of leniency cases on a CA's docket, there are also fewer non-leniency cases.

To appreciate how non-leniency enforcement can be stronger when λ is sufficiently small, consider the extreme case of $\lambda = 0$ so that leniency cases require no resources. (As results are continuous in λ, the argument will also apply when λ is sufficiently low.) Holding the cartel rate fixed at the level without a leniency program, a leniency program enhances non-leniency enforcement because there is no crowding out (due to $\lambda = 0$) *and* there are fewer non-leniency cases due to some of them being handled as leniency cases. Given that non-leniency enforcement is stronger and expected penalties are higher, the cartel rate is lower which means there are fewer non-leniency cases which serves to enhance non-leniency enforcement more. This feedback effect initiated by a leniency program - fewer cartels leads to fewer non-leniency cases which leads to stronger non-leniency enforcement which leads to fewer cartels - ultimately results in both a lower cartel rate and stronger non-leniency enforcement. This feedback effect suggests that there may be a large return to reducing λ by streamlining the handling of leniency cases.

Another circumstance in which a CA can be assured that a leniency program will lower the cartel rate is if it is introduced in a jurisdiction for which enforcement is already very weak as reflected in a low likelihood that a cartel will be even considered for prosecution. In that case, a leniency program is sure to be beneficial.

Theorem 7: There exists $\widehat{q} > 0$ such that if $q \in [0, \widehat{q}]$ then the cartel rate is weakly lower with a leniency program and if A1 holds then the cartel rate is strictly lower with a leniency program.

If non-leniency enforcement is largely absent prior to the introduction of a leniency program then a leniency program cannot have much of a crowding-out effect for the simple reason that there aren't many non-leniency cases to crowd out. Hence, if a CA is not actively engaged in enforcement prior to introducing a leniency program then a leniency program is sure to be effective in reducing the frequency of cartels.

It is worth noting that Theorem 7 is the one result that does depend on there being full leniency ($\theta = 0$). If $\theta > 0$ then, as $q \to 0$ and $\sigma\,(= qsr) \to 0$, the leniency program has no effect because no firm would use it. 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 are sufficiently low to make applying for leniency not to be in a firm's interests. The efficacy of a leniency program depends on cartel members believing there is a sufficient chance of them being caught and convicted by the CA.[22]

[22] A caveat is appropriate here because 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

5.4 Inter-Industry Variation in the Impact of a Leniency Program

Our analysis has shown how a leniency program impacts the frequency of cartels. A leniency program can lower the cartel rate by resulting in some cartels no longer forming and reducing average cartel duration for those that do form. The focus thus far has been on the aggregate cartel rate but, as the analysis of this section will reveal, a leniency program can have qualitatively distinct effects across industries.

Recall that industries vary with respect to the parameter η where a higher value of η means a higher profit increase from cheating on the collusive arrangement. A higher value for η could be due, for example, to more firms (assuming Bertrand price competition) or a higher price elasticity to the firm demand function. When η is higher, the greater incentive to deviate means that the cartel is less stable in the sense that it will internally collapse for a wider set of market conditions (specifically, $\phi(Y, \sigma, \eta)$ is decreasing in η). This property has two implications. First, industries with sufficiently high values of η are unable to cartelize (and recall that $\widehat{\eta}$ denotes the highest value for η such that a cartel forms with positive probability). Second, when cartels are able to form ($\eta \leq \widehat{\eta}$), average cartel duration is lower when η is higher. It is straightforward to show that average duration for a cartel in a type-η industry is

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

and that $\phi^*(\sigma, \eta)$ is non-increasing in η and is decreasing in η if $\phi^*(\sigma, \eta) \in (\underline{\pi}, \overline{\pi})$. Hence, average cartel duration is decreasing in η. It is useful to note that an industry type's cartel rate, $C(\sigma, \eta)$, and its average cartel duration, $CD(\sigma, \eta)$, are monotonically related:

$$C(\sigma, \eta) = \frac{\kappa(CD(\sigma, \eta) - 1)}{1 + \kappa(CD(\sigma, \eta) - 1)}.$$

Thus, in assessing how the effect of a leniency program varies across industries, we can either consider its influence on the cartel rate or on cartel duration.

Given there is inter-industry variation in the presence and duration of cartels prior to a leniency program, it is natural to examine how the impact of a leniency program varies across industries. In particular, could a leniency program make the environment *less* hospitable for collusion in some industries while making it *more* hospitable in other industries? To initially address this question, numerical analysis is conducted. Results are provided both for when the probability of conviction is linear in caseload and is a concave then convex function of caseload. Details as to parameterizations and numerical methods are in the online appendix.

will apply for leniency which is sensible for the rival only if it possesses a similar concern.

Figures 2 and 3 report the change in average cartel duration due to a leniency program for each industry type $\eta \leq \widehat{\eta}$. First note that the introduction of a leniency program reduces $\widehat{\eta}$ and thereby shrinks the range of industry types for which a cartel forms with positive probability; for example, in Figure 2(b), $\widehat{\eta}$ falls from 1.55 to 1.52. A reduction in $\widehat{\eta}$ was found for almost all parameterizations though there were a few cases in which a leniency program actually increased $\widehat{\eta}$. Second, for those industries that do cartelize with positive probability, a leniency program has a differential effect across industries depending on whether the industry produces relatively stable cartels (η is low) or unstable cartels (η is high). Specifically, the effect of a leniency program on average cartel duration (or the cartel rate) is decreasing in η so that industries that produce less stable cartels tend to experience a bigger drop in cartel duration than industries with more stable cartels. This property is apparent in Figures 2 and 3 where the change in average duration is decreasing in η, which holds as well for all other parameterizations considered. Even more than that, a leniency program can result in *longer* duration for the most stable cartels (that is, the change in duration is positive when η is low) while shutting down or shortening the duration of the least stable cartels.

Property: A leniency program generally reduces the range of markets that are able to form cartels. The effect of a leniency program on average cartel duration is decreasing in η so that markets with less stable cartels experience a bigger decline in average cartel duration. This differential effect can be so significant that a leniency program reduces average cartel duration of relatively unstable cartels and, at the same time, increases average cartel duration of relatively stable cartels.

To understand what is driving the differential effect of a leniency program across industries, recall that it is dying cartels that use the leniency program. Once market conditions are such that collusion is no longer incentive compatible, firms stop colluding and race to apply for leniency. Of course, only one firm receives leniency with the remaining firms paying full penalties. Because the leniency program then ensures conviction when the cartel dies, expected penalties are higher with a leniency program. At the same time, the flow of leniency applications can weaken non-leniency enforcement by reducing the likelihood of being prosecuted and convicted outside of the leniency program. In sum, expected penalties can be higher through the leniency program but can be lower outside of the leniency program. Which of these effects is more important depends on an industry's type. Firms in markets that support relatively unstable cartels know there is a significant chance the cartel will internally collapse which will induce a race for leniency. Thus, those cartels are especially harmed by the higher penalties coming from a leniency applicant and, therefore, they are worse off after the introduction of a leniency program. In contrast, firms in markets that support relatively stable cartels are less concerned with a race for leniency because cartel collapse is unlikely (and such a race only ensues in that event). The greater concern for a highly stable cartel is with non-leniency enforcement and, if that is weaker by virtue of the crowding-out effect of a leniency program, expected penalties are actually lower and, therefore, the environment is more hospitable for collusion.

Complementing these numerical results, we can prove that the least stable cartels are harmed by a leniency program when non-leniency enforcement is not weakened, and the most stable cartels are benefitted by a leniency program when non-leniency enforcement is weakened. Theorem 8 shows that if a market produces sufficiently stable cartels then the impact of a leniency program is determined by how it influences non-leniency enforcement. If the introduction of a leniency program strengthens non-leniency enforcement then the cartel rate (or average cartel duration) for highly stable cartels declines, while if it weakens non-leniency enforcement then the cartel rate (or average cartel duration) rises. These highly stable cartels are not concerned with the higher penalties coming from a race for leniency - because a race is unlikely for those cartels - and instead are concerned with whether they are more or less likely to be prosecuted and convicted outside of the leniency program.

Theorem 8: Assume $\underline{\eta} = 1$ and $C_{NL}\left(\sigma_{NL}^*, \eta\right), C_L\left(\sigma_L^*, \eta\right) > 0$ then

$$\lim_{\eta \to \underline{\eta}} \left[C_L\left(\sigma_L^*, \eta\right) - C_{NL}\left(\sigma_{NL}^*, \eta\right)\right] = \frac{\kappa\left(\sigma_{NL}^* - \sigma_L^*\right)}{\left[1-(1-\kappa)\left(1-\sigma_L^*\right)\right]\left[1-(1-\kappa)\left(1-\sigma_{NL}^*\right)\right]}.$$

Hence, for η close to one, $C_L\left(\sigma_L^*, \eta\right) < (>)\, C_{NL}\left(\sigma_{NL}^*, \eta\right)$ if and only if $\sigma_L^* > (<)\, \sigma_{NL}^*$.

The next result shows that, unless non-leniency enforcement is weakened, a leniency program is sure to destabilize the least stable cartels. More specifically, industries for which $\eta \in \left(\widehat{\eta}_L\left(\sigma_L^*\right), \widehat{\eta}_{NL}\left(\sigma_{NL}^*\right)\right]$ were able to collude in the absence of a leniency program but are not able to do so with a leniency program.

Theorem 9: If $\sigma_{NL}^*, \sigma_L^* \in (0, \omega)$, $\sigma_L^* \geq \sigma_{NL}^*$, and $\widehat{\eta}_{NL}\left(\sigma_{NL}^*\right) \in \left(\underline{\eta}, \overline{\eta}\right)$ then $\widehat{\eta}_L\left(\sigma_L^*\right) < \widehat{\eta}_{NL}\left(\sigma_{NL}^*\right)$.

In sum, the institution of a leniency program can raise expected penalties and, as a result, shorten cartel duration (or prevents cartels from forming at all) in industries for which collusion is least stable, while it can actually lengthen cartel duration in industries for which collusion is most stable because non-leniency enforcement is weaker. Thus, this theory predicts that a leniency program can result in fewer cartels forming but those that form last a longer time.

Finally, these results could be the basis for a test of whether a leniency program is weakening non-leniency enforcement which is a crucial condition for a leniency program to raise the cartel rate. A candidate marker is that a leniency program results in the most stable cartels having longer duration and the least stable cartels having shorter duration. For example, consider the change in average duration between before and after instituting a leniency program for the $X\%$ of cartels with the longest duration and the $X\%$ of cartels with shortest duration, for some $X \leq 50$. If the former rises and the latter falls then this is consistent with weaker non-leniency enforcement. Proper development of this test requires further work, however, because it is stated in terms of the duration of cartels, while the data will be the duration of *discovered* cartels. Given a leniency program affects the discovery process, if the chances of discovery are correlated with the change in duration then the change in duration for discovered cartels will be a biased measure of the change in durations for cartels (see Harrington and Wei [2014] for an examination of selection bias with regards to discovered cartel data). A second source of possible bias is that the least stable cartels may no longer form in which case they are not present in the post-leniency data. The framework of this paper can be used to take account of those biases and develop a proper test of whether a leniency program is weakening non-leniency enforcement.

5.5 Robustness

In concluding this section, let us investigate the robustness of the result that a leniency program can raise the cartel rate (Theorem 4) with respect to the assumption that the CA pursues all leniency cases. For this purpose, let x denote the fraction of leniency cases that the CA takes on and consider $x < 1$. If a firm applies for leniency and the CA chooses not to accept the application then the CA cannot prosecute the case (or, if it did, then the court would annul any penalties.)[23] Recall that, when all members apply for leniency upon a cartel's death, the expected penalty to a cartel member is ωF. If now only a fraction x of leniency applications are accepted then the expected penalty is reduced to $x\omega F + (1 - x) \times 0 = x\omega F$. This creates a serious enforcement problem when x is low. If x is such that $x\omega F < \sigma F$ then expected penalties are *lower* under a leniency program (in contrast to our assumption that they are higher) in which case it becomes *easier* for a leniency program to raise the cartel rate. Furthermore, if x is set too low then active cartels will also apply for leniency because they know there is a good chance that the application will be denied which then prevents the CA from prosecuting them. Thus, there are compelling reasons for the CA not to pursue too few leniency cases and thereby keep x reasonably high.

Next suppose the CA pursues a large fraction of leniency cases but not all of them, and further suppose $x\omega F > \sigma F$ so that expected penalties are higher with a leniency program. Where this changes the analysis is in the determination

[23] Email from Kevin Coates (head of the cartels directorate unit of the European Commission) to Joseph Harrington, July 29, 2014.

of the equilibrium value of σ. From equation (12) for the case of full leniency ($\theta = 0$), σ^* is now the fixed point to

$$\sigma = qrp\left(\lambda x L\left(\sigma\right) + qr\left(C\left(\sigma\right) - L\left(\sigma\right)\right)\right).$$

$L(\sigma)$ is the mass of leniency applications from which the CA pursues $xL(\sigma)$ leniency cases. Out of the $C(\sigma) - L(\sigma)$ cartels who did not apply for leniency, the CA discovers a fraction q of them and prosecutes a fraction r of those cases. It can be shown that the condition in Theorem 4 is now $\lambda x > qr$ instead of $\lambda > qr$. For example, if $\lambda \simeq 1$ so leniency cases take roughly as many resources as non-leniency case then a leniency program raises the cartel rate when penalties are sufficiently weak and $qr/x < 1$. The latter condition is satisfied if the rate at which the CA takes on leniency cases is at least as great as the rate at which it takes on non-leniency cases ($x \geq r$) which seems reasonable given that the success rate for a leniency case exceeds that for a non-leniency case. Thus, as long as the CA is sufficiently attracted to leniency cases relative to non-leniency cases then the cartel rate will be higher with a leniency program when penalties are sufficiently weak and the resource savings in prosecuting a leniency case are sufficiently small.

6 Concluding Remarks: Policy Implications and Future Research Directions

> U.S. Senator Bill Blumenthal speaking to William Baer, Assistant Attorney General, Antitrust Division, U.S. Department of Justice: The fear or apprehension – in other words, the deterrent effect of past prosecution – is what drives the leniency program at the end of the day. My concern is that most of the cases that are brought today are ... generated exclusively from firms that decided to come forward and seek a leniency application . . . . I'm worried that the success of the leniency program combined with budget constraints that your Division faces will, in effect, give you incentives to pursue only the companies that come forward . . . [A]s I know from personal experience, some of the most egregious and harmful of the cartels may have nobody coming forward. [U.S. Senate Hearing on "Cartel Prosecution: Stopping Price Fixers and Protecting Consumers' Subcommittee on Antitrust, Competition Policy and Consumer Rights, November 14, 2013]

> [T]he Division should be mindful that relying too heavily on leniency may be detrimental to its overall goal of decreasing harmful cartel activity ... [A] regime wherein cartelists may fear being exposed by their co-conspirators in exchange for leniency, but where they face no real danger of otherwise being detected, is lopsided and thus less

effective both as a detector of and a deterrent to bad behavior than if resources were more evenly allocated between deterrence and detection.' Megan Dixon, Ethan Kate, and Janet McDavid - Hogan Lovells LLP, *CPI Antitrust Chronicle*, February 2014 (2)

The sentiment expressed by Senator Blumenthal and attorneys at Hogan Lovells LLP is a common one. While recognizing that a leniency program is an immensely valuable tool in the arsenal of a competition authority, concerns arise when it is predominantly the only tool. Underlying this concern is that leniency applications and convictions - of which the U.S. Department of Justice and the European Commission have in ample supply - are not a reassuring measure of performance. The real issue is whether fewer cartels are forming and persisting.

The concerns expressed by both practitioners and scholars were investigated in this paper by constructing a framework within which to examine how a leniency program influences the intensity of enforcement through non-leniency means and how both forms of enforcement affect the frequency of cartels. Holding non-leniency enforcement fixed, we find that a leniency program generally lowers the cartel rate, which is consistent with previous theoretical and experimental research. However, when non-leniency enforcement is endogenized, a leniency program can either lower or raise the cartel rate. Whether there are more or less cartels depends on the extent to which leniency applications shift competition authority resources away from pursuing cases without a leniency application to cases with a leniency application. In particular, introducing a leniency program into an environment in which penalties are low and leniency cases take up comparable resources to that of non-leniency cases is predicted to result in a higher frequency of cartels.

Our analysis suggests some feasible policies that can make it more likely that the introduction of a leniency program will have the desired effect of reducing the presence of cartels in an economy. A leniency program shifts resources from prosecuting active cartels to inactive cartels which is counter-productive in terms of shutting down cartels. Prosecution of dead cartels can still be beneficial, however, when it enhances deterrence. Indeed, a primary appeal of a leniency program is that it will cause cartel members to believe there is a higher probability that they'll end up paying penalties, but this will substantively deter cartel formation only if penalties are sufficiently severe. It is then important that the institution of a leniency program be done in conjunction with appropriately high penalties. Failure to do so can mean a leniency program is not just ineffective but counter-productive. A second key complementary policy to a leniency program is a competition authority's budget. With firms coming to the competition authority for leniency - rather than a competition authority having to actively look for cases - there may be a temptation to reduce the resources of a competition authority. This would be a mistake. If resources are constrained then the attention given to leniency cases will have a crowding-out effect on non-leniency enforcement which can allow the more stable cartels to avoid prosecution longer and this can raise the presence of cartels. A competition authority's budget and a leniency program should be thought of as complements - not substitutes - and, accordingly, budgets should be expanded and processes should be streamlined to more effectively handle leniency cases so as to maintain a strong enforcement presence for cartels who do not apply for leniency.

While the focus in this paper has been on assessing the impact of a corporate leniency program, the framework is flexible enough to evaluate other competition policies. For example, assessing the impact of screening for cartels, which can be represented as an increase in the probability that a cartel is discovered q (see Harrington [2007] for an argument for screening). Or endowing the competition authority with some amount of resources and considering their allocation across discovery (reducing q), prosecution (increasing the probability of conviction s), and penalization (increasing penalties γ). The model can also be enriched to give the competition authority more powers by, for example, allowing the penalty to depend on the actual duration of the cartel (rather than average duration) or allowing the competition authority to observe an industry's type η and then select cases on those grounds; in particular, it would want to take cases with lower η because cartels tend to be more stable in those markets and thus shutting them down is more important.

Perhaps the most intriguing extension is modelling the competition authority's objective and exploring optimal policy design when the competition authority is motivated by career concerns. If we think that the members of the competition authority act to maximize some observable measure of performance then, given that the cartel rate is not directly observed, it isn't clear there is an incentive scheme that will induce it to minimize the cartel rate. A more reasonable objective might be for it to maximize the number of convictions or the total amount of fines. In that case, the ultimate effect of a policy change can depend on how it is implemented by the competition authority. Better understanding the strategic behavior of a competition authority and how it impacts the efficacy of competition policy is an important avenue for future research.

References

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

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

[3] Blatter, Marc, Winand Emons, and Silvio Sticher, "Optimal Leniency Programs when Firms Have Cumulative and Asymmetric Evidence,' Universität Bern, July 2014.

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

[5] Borrell, Joan-Ramon, Juan Luis Jiménez, and Carmen García. 2014. "Evaluating Antitrust Leniency Programs,' *Journal of Competition Law and Economics*, 10, 107-136.

[6] Bos, Iwan and Frederick Wandschneider. 2013. "A Note on Cartel Ringleaders and the Corporate Leniency Program,' *Applied Economics Letters*, 20, 1100-1103.

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

[8] Bundeskartellamt. July 2010. "Effective Cartel Prosecution: Benefits for the Economy and Consumers."

[9] Chen, Joe and Joseph E. Harrington, Jr. 2007. "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.

[10] Chen, Zhijun and Patrick Rey. 2013. "On the Design of Leniency Programs,' *Journal of Law and Economics*, 56, 917-957.

[11] Choi, Yun Jeong. 2011. "An Empirical Analysis of the Corporate Leniency Program in Korea: Its Amendments in 2005,' *Korean Journal of Economics*, 18, 255-281.

[12] Choi, Jay Pil and Heiko Gerlach. 2012. "Global Cartels, Leniency Programs and International Antitrust Cooperation,' *International Journal of Industrial Organization,* 30, 528–540.

[13] Dijkstra, Peter T., Marco A. Haan, and Lambert Schoonbeek. 2011. "Leniency Programs and the Design of Antitrust: Experimental Evidence with Unrestricted Communication,' University of Groningen.

[14] Dong, Ailin, Massimo Massa, and Alminas Žaldokas. March 2014. "Busted! Now What? Effects of Cartel Enforcement on Firm Values and Policies,' INSEAD.

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

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

[17] Gärtner, Dennis. 2013. "Corporate Leniency in a Dynamic World: The Preemptive Push of an Uncertain Future,' University of Bonn, October.

[18] Ghosal, Vivek and Joseph Gallo. 2001. "The Cyclical Behavior of the Department of Justice's Antitrust Enforcement Activity,' *International Journal of Industrial Organization,* 19, 27-54.

[19] Hamaguchi, Yasuyo, Toshiji Kawagoe, and Aiko Shibata. 2009. "Group Size Effects on Cartel Formation and the Enforcement Power of Leniency Programs,' *International Journal of Industrial Organization*, 27, 145–165.

[20] Hammond, Scott D. 2005. "An Update of the Antitrust Division's Criminal Enforcement Program," ABA Section of Antitrust Law, Cartel Enforcement Roundtable, November 16.

[21] Harrington, Joseph E., Jr. 2007. "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.

[22] Harrington, Joseph E., Jr. 2008. "Optimal Corporate Leniency Programs,' *Journal of Industrial Economics*, 56, 215-246.

[23] Harrington, Joseph E., Jr. 2012. "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.

[24] Harrington, Joseph E., Jr. 2013. "Corporate Leniency Programs when Firms have Private Information: The Push of Prosecution and the Pull of Preemption,' *Journal of Industrial Economics*, 61, 1-27.

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

[26] Harrington, Joseph E., Jr. and Yanhao Wei. 2014. "What Can the Duration of Discovered Cartels Tell Us About the Duration of Cartels?,' The Wharton School, University of Pennsylvania, July.

[27] Hinloopen, Jeroen and Adriaan R. Soetevent. 2008. "Laboratory Evidence on the Effectiveness of Corporate Leniency Programs,' *RAND Journal of Economics*, 39, 607-616.

[28] Houba, Harold, Evgenia Motchenkova, and Quan Wen. 2009. "The Effects of Leniency on Maximal Cartel Pricing,' TILEC Discussion Paper No. 2009-038, October.

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

[30] Klein, Gordon. 2010. "Cartel Destabilization and Leniency Programs - Empirical Evidence,' ZEW Discussion Paper No. 10-107.

[31] Kwoka, John. 1999. "Commitment to Competition: An Assessment of Antitrust Agency Budgets since 1970,' *Review of Industrial Organization,* 14, 295-302.

[32] Lefouili, Yassine and Catherine Roux. 2012. "Leniency Programs for Multimarket Firms: The Effect of Amnesty Plus on Cartel Formation,' *International Journal of Industrial Organization*, 30, 624-640.

[33] Marshall, Robert C., Leslie M. Marx, and Claudio Mezzetti. 2013. "Antitrust Leniency with Multi-Product Colluders,' Penn State University, May.

[34] Marx, Leslie M. and Claudio Mezzetti. 2014. "Effects of Antitrust Leniency on Concealment Effort by Colluding Firms,' Duke University, Fuqua School of Business, April (*Journal of Antitrust Enforcement*, forthcoming)..

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

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

[37] Motchenkova, Evguenia. 2004. "Effects of Leniency Programs on Cartel Stability,' Center for Economic Research Discussion Paper No. 2004-98, Tilburg University.

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

[39] Report to the 7th ICN Annual Conference, Kyoto. 2008. "Setting of Fines for Cartels in ICN Jurisdictions," April.

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

[41] Koh, Sae Ran and Jinook Jeong. 2013. "Leniency Program in Korea and Its Effectiveness,' *Journal of Competition Law and Economics*, 9, 1-23.

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

[43] Sauvagnat, Julien. 2012. "Prosecution and Leniency Programs: The Role of Bluffing in Opening Investigations,' Toulouse School of Economics, (*Journal of Industrial Economics*, forthcoming).

[44] Silbye, Frederik. 2010. *Topics in Competition Policy: Cartels, Leniency, and Price Discrimination*, PhD Thesis, University of Copenhagen, August.

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

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

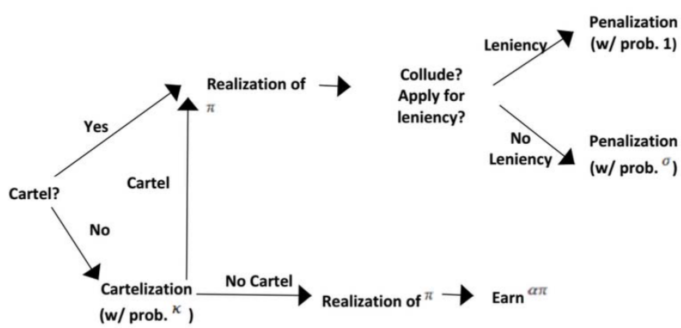
[47] Stephan, Andreas. 2008. "An Empirical Assessment of the European Leniency Notice,' *Journal of Competition Law and Economics*, 5, 537-561.

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

Table 1

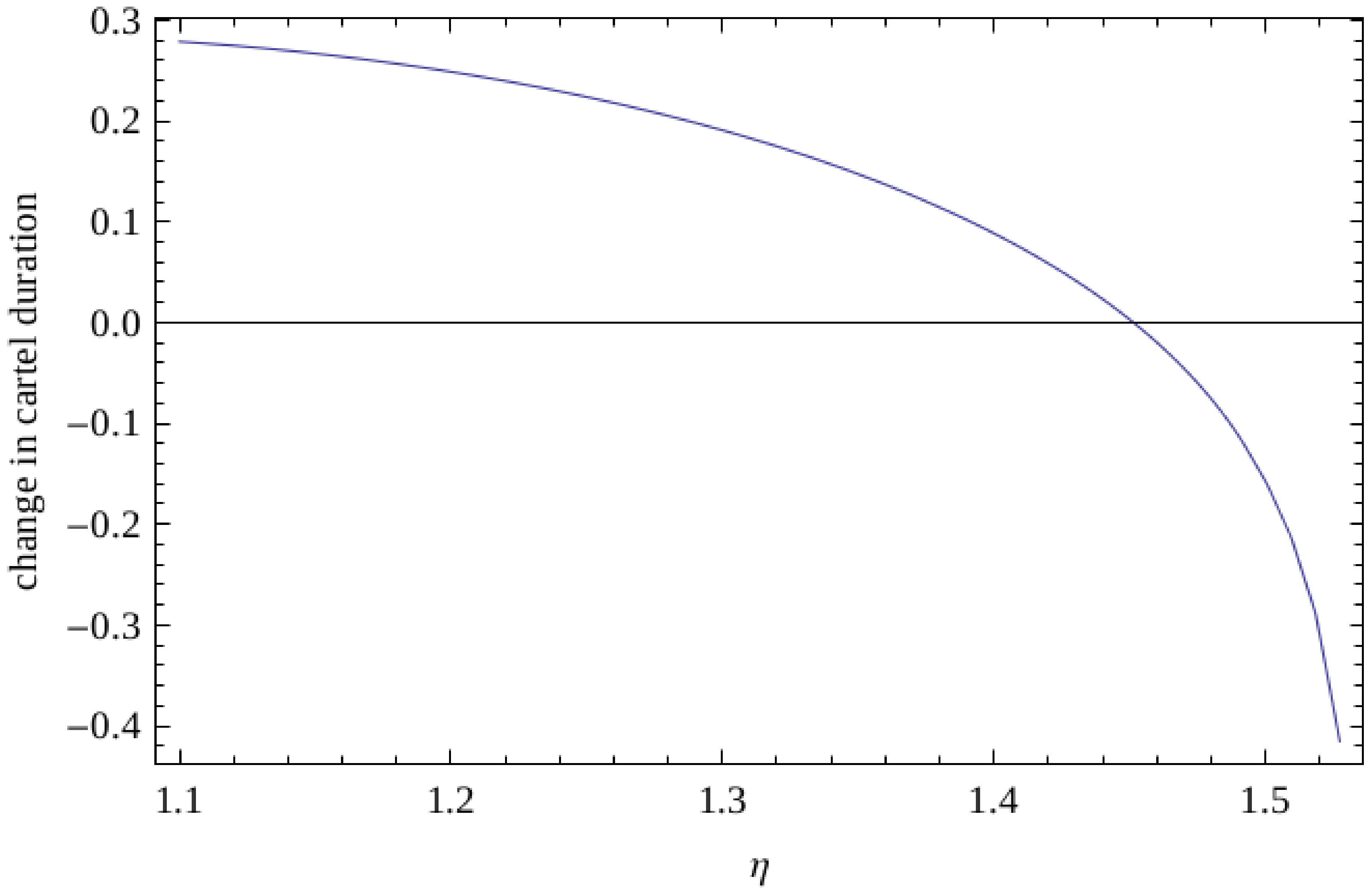
Own action	All other firms' actions	Own profit
collude	*collude*	π
compete	*collude*	$\eta\pi$
compete	*compete*	$\alpha\pi$

Figure 1



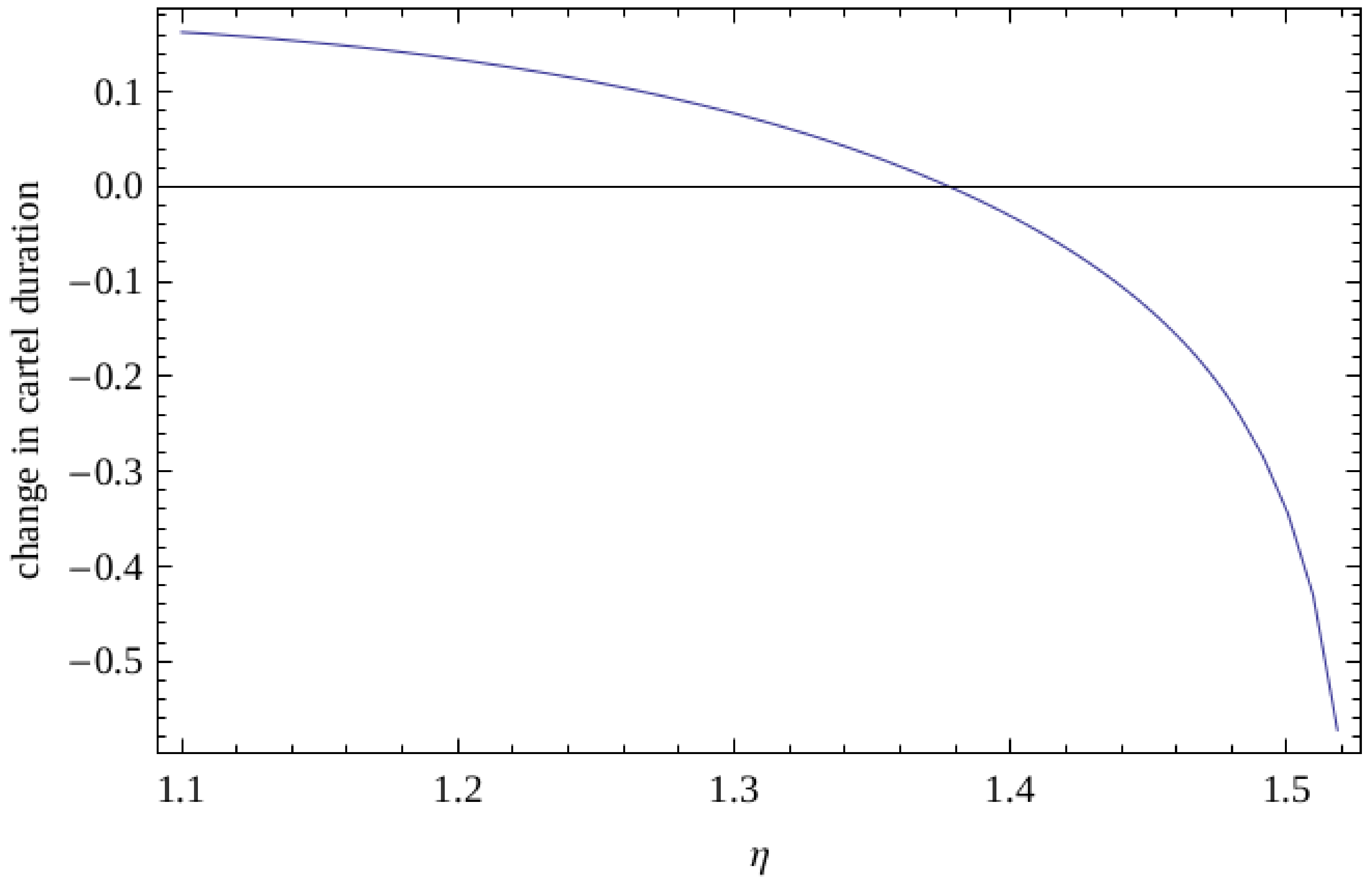
(a) γ = .7 and λ = 1.0

$\hat{\eta}_L = 1.5272, \hat{\eta}_{NL} = 1.5539$



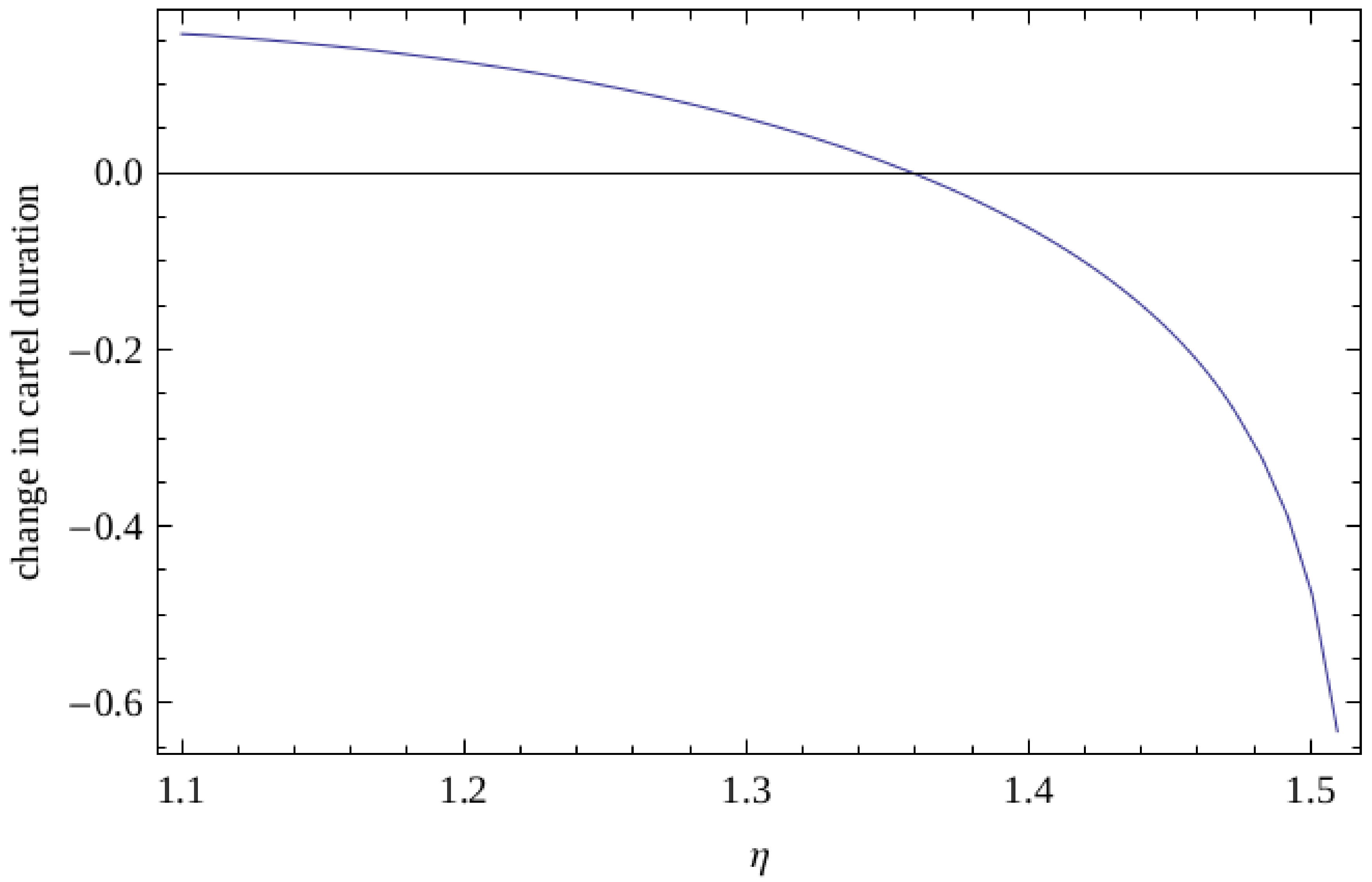
(b) γ = .8 and λ = 1.0

$\hat{\eta}_L = 1.5183, \hat{\eta}_{NL} = 1.5539$

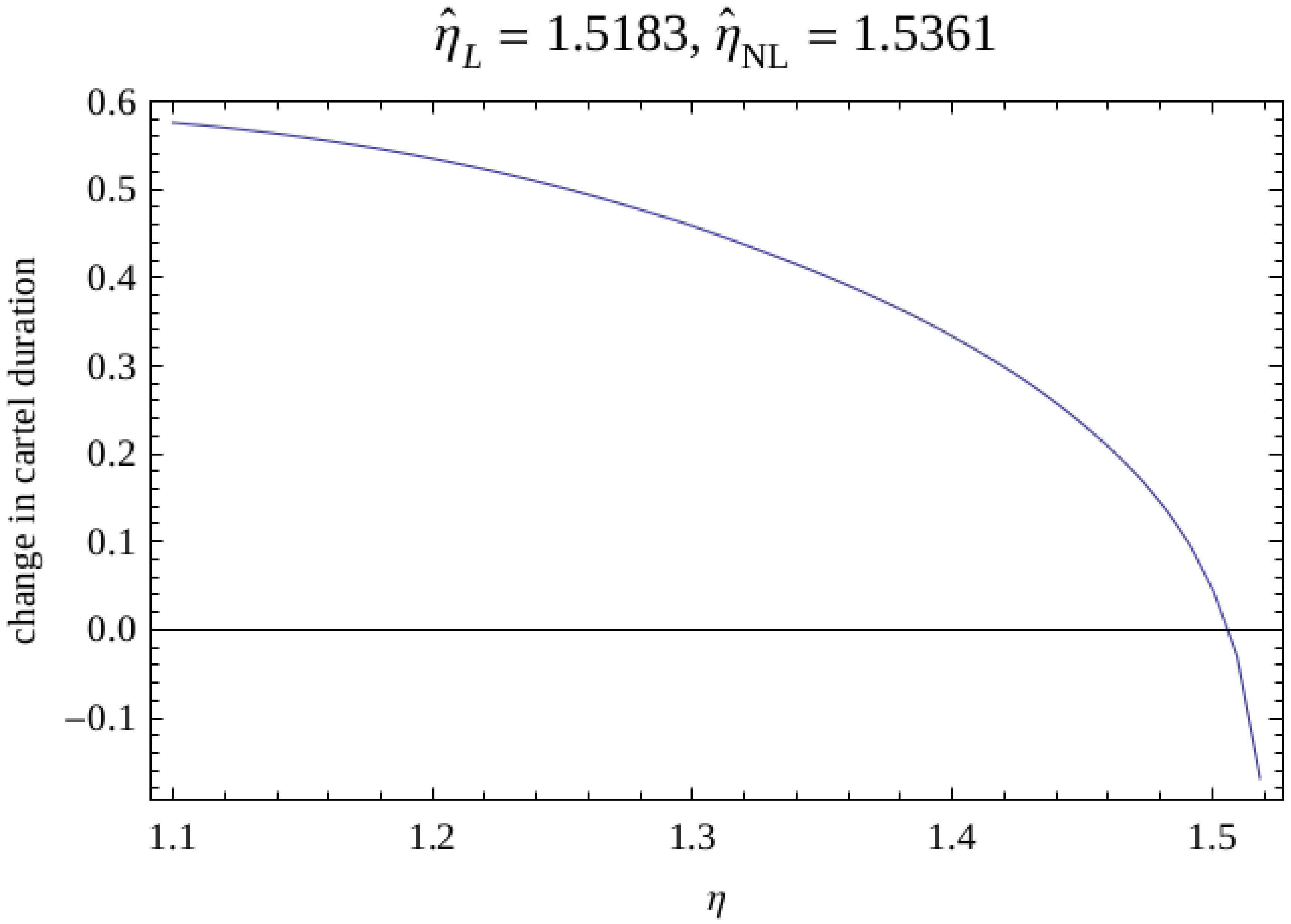


(c) γ = .9 and λ = 1.0

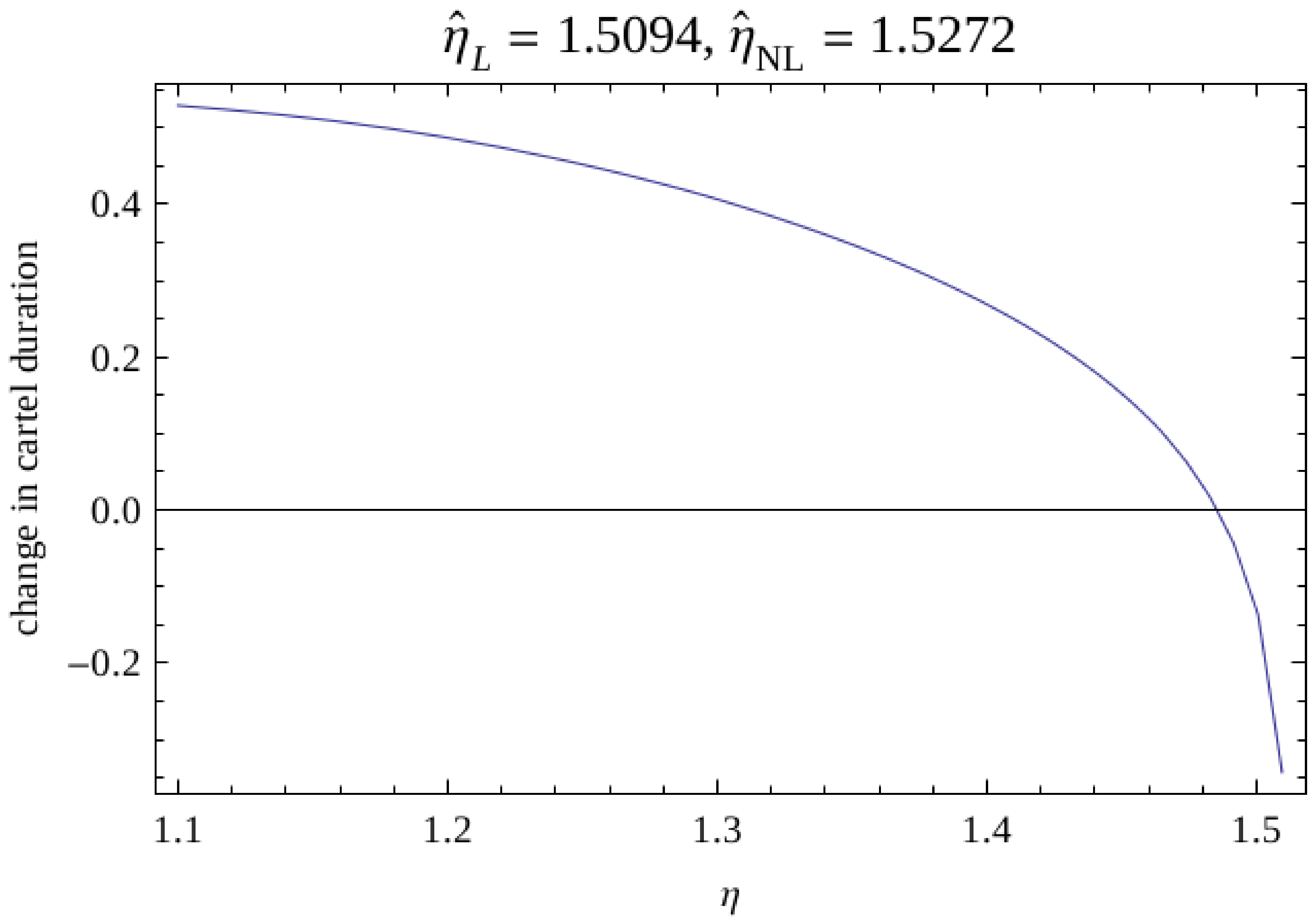
$\hat{\eta}_L = 1.5094, \hat{\eta}_{NL} = 1.5450$



(a) γ = .7 and λ = 1.0



(b) γ = .8 and λ = 1.0



(c) γ = .9 and λ = 1.0