An Optimal Rule for Patent Damages Under Sequential Innovation

Chen, Yongmin and Sappington, david

University of Colorado Boulder, University of Florida

30 August 2016

An Optimal Rule for Patent Damages Under Sequential Innovation

by

Yongmin Chen* and David E. M. Sappington**

August 2016

Abstract: We analyze the optimal design of damages for patent infringement in settings where the patent of an initial innovator may be infringed by a follow-on innovator. We consider damage rules that are linear combinations of the popular "lost profit" (LP) and "unjust enrichment" (UE) rules, coupled with a lump-sum transfer between the innovators. We identify conditions under which a linear rule can induce the socially optimal levels of sequential innovation and the optimal allocation of industry output. We also show that, despite its simplicity, the optimal linear rule achieves the highest welfare among all rules that ensure a balanced budget for the industry, and often secures substantially more welfare than either the LP rule or the UE rule.

Keywords: Patents, sequential innovation, infringement damages, linear rules for patent damages.

* Department of Economics, University of Colorado; Yongmin.Chen@colorado.edu

** Department of Economics, University of Florida; sapping@ufl.edu

We gratefully acknowledge helpful comments from Ted Bergstrom, Michael Meurer, Marius Schwartz, Yair Tauman, John Turner, and seminar participants at Fudan University, Queens University, the State University of New York at Stony Brook, the University of California at Santa Barbara, the University of Georgia, the University of Groningen, the 2016 Workshop on IO and Competition Policy at the Shanghai University of Finance and Economics, and the 2016 Workshop on Industrial Organization and Management Strategy at the Hong Kong University of Science and Technology.

1 Introduction

Innovation is a key driver of economic growth and prosperity. To encourage innovation, successful innovators often are awarded patents for their inventions. A patent grants an innovator exclusive rights to her invention for a specified period of time. An extensive literature analyzes optimal patent protection, focusing on issues such as the optimal strength and breadth of patents.[1] An important, but less developed, literature studies financial penalties ("damages") for patent infringement. To date, this literature has primarily analyzed the performance of individual damage rules that are employed in practice, including the *lost profit* (LP) rule and the *unjust enrichment* (UE) rule.[2] In contrast, the purpose of this paper is to analyze the optimal design of patent damage rules under sequential innovation, where an initial innovator's patent may be infringed by a follow-on innovator whose differentiated product, possibly of higher quality, expands market demand.

Sequential innovation is important to consider because it drives progress in many modern-day industries. To illustrate, today's smartphones are estimated to embody innovations protected by as many as 250,000 patents that have been developed sequentially (Sparapani, 2015).[3] The design of damages for patent infringement is particularly subtle in the presence of sequential innovation because, although stringent damage rules can encourage early innovation, they may discourage subsequent innovation, especially when uncertainty prevails about whether follow-on innovations infringe earlier patents.[4]

---

[1] Early studies of optimal patent protection include Nordhaus (1969), Gilbert and Shapiro (1990), Klemperer (1990), and Scotchmer (1991)

[2] Anton and Yao (2007) examine the performance of the LP rule. Shankerman and Scotchmer (2001), Choi (2009), and Henry and Turner (2010) examine the performance of both the LP and the UE rules. (These rules are described below.) Choi and Henry and Turner also analyze the performance of the *reasonable royalty* damage rule, which is discussed in section 6.

[3] It has been observed that modern computing technologies like smart phones "tightly and efficiently integrate the engineering of other companies and other earlier inventors, and are enhanced with new functions and features to attract consumer interest and drive demand" (Sparapani, 2015). Similarly, discoveries regarding gene sequencing are valuable inputs in follow-on scientific research and in commercial applications (Sampat and Williams, 2015). Murray et al. (2016) examine the effects of affording academic researchers expanded access to newly discovered information about genetically engineered (transgenic) mice. The authors report that expanded access increases follow-on discoveries by new researchers and promotes diverse follow-on research methodologies without reducing the rate of innovation.

[4] Green and Scotchmer (1995) examine the patent length and division of surplus required to induce efficient

We consider a model in which innovation is not certain due to stochastic variation in innovation costs. Patent protection also is uncertain in our model, as it is in practice.[5] Hundreds of thousands of patents are granted annually, and patent descriptions can be vague and incomplete.[6] Therefore, in practice, it is often difficult to discern whether an innovation infringes an existing patent.[7] The parameter $\lambda \in (0, 1]$ in our model denotes the probability that the patent of an initial innovator, firm 1, is infringed by the differentiated product of a follow-on innovator, firm 2. The value of λ can be viewed as a measure of the strength of patent protection (e.g., Choi, 1998; and Farrell and Shapiro, 2008).

We consider damage rules that are linear combinations of the LP rule and the UE rule, coupled with a lump-sum transfer between the innovators. The LP rule requires the infringer to compensate the patent holder for the reduction in profit the latter suffers due to the infringement.[8] The UE rule requires the infringer to deliver its realized profit to the patent holder.[9] Under the linear rules that we analyze, if firm 2 is found to have infringed firm 1's patent, firm 2 must deliver a damage payment (D) to firm 1 that has three components: a

sequential innovation. We extend their work in part by explicitly modeling competition between innovators in the presence of uncertainty about the applicability of existing patents.

[5] Anton and Yao (2007), Choi (2009), and Henry and Turner (2010) also analyze probabilistic patent enforcement.

[6] See, for example, Choi (1998), Lemley and Shapiro (2005), Bessen and Meurer (2008), and Farrell and Shapiro (2008).

[7] The recent protracted patent infringement ligation between Apple and Samsung is a case in point (e.g., Vascellaro, 2012). In addition, Lemley and Shapiro (2005) report that the U.S. Patent and Trademark Office issues nearly 200,000 patents annually, so the time that a patent officer can devote to assessing the merits of any individual patent application is limited. Consequently, even after a patent is issued, its validity may be successfully contested in court. Uncertainty about prevailing patent protection also can complicate the licensing of innovations. See Kamien and Tauman (1986), Katz and Shapiro (1986), and Scotchmer (1991), for example, for analyses of the licensing of innovations.

[8] U.S. patent law stipulates that the damage penalty for patent infringement must be "adequate to compensate for the infringement, but in no event less than a reasonable royalty for the use made of the invention by the infringer, together with interest and costs as fixed by the court." Courts have interpreted this stipulation to require "an award of lost profits, or other compensatory damages, where the patentee can prove, and elects to prove, such damages" (Frank and DeFranco, 2000-2001, p. 281).

[9] The UE rule is sometimes employed when a design patent is infringed. The Supreme Court of the United States recently agreed to review a Federal Appeals Court decision that Apple is entitled to all of the profit that Samsung earned from its sales of smartphones because Samsung "infringed three Apple design patents relating to a portion of the iPhone's outer shell and a single graphical-user-interface screen" (Amici Curiae Brief, 2016).

lump sum monetary payment (m), a fraction (d_1) of the amount by which firm 2's operation reduces firm 1's profit, and a fraction (d_2) of firm 2's profit. Thus, linear rules generalize the LP rule and the UE rule, including the former (with $d_1 = 1$ and $m = d_2 = 0$) and the latter (with $d_2 = 1$ and $m = d_1 = 0$) as special cases.[10]

We find that although linear rules are relatively simple in nature, an optimally-designed linear rule secures the highest welfare among all balanced damage rules (i.e., rules in which all payments are internal to the industry). The optimality of the linear rule in this sense reflects in part its links to both the (lost) profit of the initial innovator and the (realized) profit of the follow-on innovator. These links enable an optimally-designed linear rule to secure desired levels of industry profit, distribute the profit to foster innovation by both firms, and ensure desired allocations of industry output.

We also show that the optimal linear rule typically differs from both the LP rule and the UE rule. Furthermore, the optimal linear rule often can secure a substantial increase in welfare relative to both the LP rule and the UE rule. In addition, when the maximum feasible level of industry profit is sufficiently large relative to innovation costs, the optimal linear rule can ensure the first-best outcome, under which each firm innovates if and only if its innovation enhances welfare and industry output is allocated between the producers so as to maximize welfare. When the optimal linear rule achieves the first-best outcome, it resembles the LP rule more closely than the UE rule (so $d_1 > d_2$) when consumers value the product of the initial innovator relatively highly, while it resembles the UE rule more closely (so $d_2 > d_1$) when consumers value firm 2's product relatively highly.[11] We also show how the optimal linear rule can limit the distortions in innovation incentives and output allocation when the first-best outcome is not attainable.

[10] Linear rules also encompass profit sharing rules, which require a patent infringer to pay the patent holder a portion of the profit the infringer secures in the marketplace.

[11] As we demonstrate below, the optimal linear rule serves to shift equilibrium industry output toward the product that consumers value most highly. The linear rule affects the equilibrium allocation of industry output by inducing the firms to partially internalize each other's profit, which influences their pricing strategies.

The welfare-maximizing linear rule is optimal among balanced rules regardless of the strength of protection for firm 1's patent (i.e., for any given $\lambda > 0$).[12] This finding suggests that although the appropriate strength and scope of patents can help to foster efficient levels of sequential innovation, the design of damages for patent infringement may play a particularly important role in this regard.

The ensuing analysis proceeds as follows. Section 2 presents the model. Section 3 records equilibrium market outcomes. Section 4 characterizes and explains the key features of the optimal linear rule. Section 5 demonstrates that the welfare-maximizing linear rule is optimal among balanced damage rules. Section 5 also illustrates the (often substantial) welfare gains that the optimal linear rule can secure relative to popular damage rules. Section 6 summarizes our findings, considers extensions of our analysis, and suggests directions for further research. Appendix A presents the proofs of all formal conclusions. Appendix B further illustrates the optimal linear rule and the welfare gains it can secure.

2 The Model

We consider the interaction between two firms. Firm 1 has the potential to innovate first and secure a patent on its product by incurring innovation cost k_1. Firm 2 has the potential to innovate subsequently by incurring k_2, but only if firm 1 has innovated successfully.[13] These innovation costs are the realizations of independent random variables, with continuous and strictly increasing distribution functions $F(k_1)$ and $G(k_2)$, respectively, on support $\left[0, \bar{k}_i\right]$ for $i = 1, 2$. We assume that innovation costs are the only costs the firms incur.[14]

[12] It follows that the optimal linear rule also dominates *ex ante* efficient patent licensing, as we demonstrate in section 6.

[13] Thus, we follow Chang (1995) in abstracting from stochastic innovation and from the possibility of simultaneous research and development activity by multiple firms. We extend Chang's work in part by considering product differentiation and uncertainty about the applicability of existing patents. Bessen and Maskin (2009) allow for stochastic, simultaneous research and development by multiple firms. They show that patent protection can reduce welfare in settings where the research activities of distinct firms are complementary.

[14] Thus, we adopt Green and Scotchmer (1995)'s "idea" approach to modeling innovation, assuming that innovators naturally acquire ideas about new products and an innovator must incur a specified cost to implement the innovation.

A mass of Q potential consumers are distributed uniformly on a line segment of length Q. If firm 1 innovates, it locates at point 0 on the line segment, and $Q = 1$ if firm 1 is the only innovator. If firm 2 innovates, it locates at point $Q = L \geq 1$ on the line segment, where L is an exogenous parameter that captures the extent to which firm 2's innovation expands the market. The product supplied by firm i delivers value v_i to consumers for $i = 1, 2$. The net utility a consumer derives from traveling distance s to purchase a unit of the product from firm i at price p_i is $v_i - p_i - s$. Each consumer purchases at most one unit of the product, and purchases from the supplier that offers the highest nonnegative net utility. This formulation admits both monopoly and duopoly industry structures. The duopoly structure extends the standard Hotelling model by allowing the follow-on innovation of firm 2 to benefit consumers not only through (standard) horizontal product differentiation,[15] but also through quality improvement (when $v_2 > v_1$) and market expansion (when $L > 1$).[16]

To capture relevant uncertainty about whether firm 2's product infringes firm 1's patent, we let $\lambda \in (0, 1]$ denote the probability that, after serving consumers, firm 2 is ultimately found to have infringed firm 1's patent.[17] In this event, firm 2 is obligated to pay firm 1 the amount $D \equiv m + d_1 R_1 + d_2 \pi_2$, where R_1 is the amount by which firm 1's profit is reduced by firm 2's operation, π_2 is firm 2's profit, and m is a lump sum monetary payment. The policy instruments, m, $d_1 \geq 0$, and $d_2 \geq 0$, are chosen to maximize expected welfare (the sum of consumer and producer surplus). We will call the damage rule associated with damage payment D the *linear rule*.

The timing in the model is as follows. First, λ is determined exogenously. Then d_1, d_2,

[15] Henry and Turner (2010) consider price competition in a standard Hotelling model. Anton and Yao (2007) and Choi (2009) analyze (Cournot) quantity competition with a homogeneous product.

[16] As we explain in section 6, our primary findings persist in more general models of competition with product differentiation. We focus on this Hotelling model with potential market expansion because the closed-form solutions it admits facilitate the demonstration of our findings.

[17] We take as given any actions firm 2 might undertake to avoid patent infringement. Gallini (1992) demonstrates that long patent durations can reduce welfare by encouraging firms to "invent around" the patents of early innovators. Zhang and Hylton (2015) analyze the optimal design of infringement penalties in a setting where the potential infringer can undertake unobserved actions to limit the loss the patent holder suffers in the event of infringement.

and m are set to maximize expected welfare. Next, firm 1 privately learns the realization of k_1 and decides whether to innovate. If firm 1 decides not to innovate, the game ends. If firm 1 innovates by incurring cost k_1, then firm 2 privately learns the realization of k_2 and decides whether to incur this cost in order to secure the follow-on innovation. If firm 2 innovates, it subsequently competes against firm 1 for the patronage of consumers on $[\,0,\,L\,]$. If firm 2 chooses not to innovate, then firm 1 acts as a monopolist and serves consumers on $[\,0,\,1\,]$. When both firms compete in the marketplace, they set their prices simultaneously and non-cooperatively. After consumers have been served, it becomes known whether firm 2 has infringed firm 1's patent. If firm 2 is found to have infringed 1's patent, firm 2 makes the requisite payment to firm 1.

We assume throughout the ensuing analysis that consumer valuations of the two products are not too disparate, and are large relative to transportation costs. This assumption helps to ensure that all consumers purchase a unit of the product and that, under duopoly competition, both firms serve customers in equilibrium.[18]

Assumption 1. $\min\{v_1, v_2\} > 2\,L$ and $|\,\Delta\,| < L$, where $\Delta \equiv v_2 - v_1$.

3 Market Equilibrium

Having identified the key elements of the model, we now characterize equilibrium outcomes. To begin, consider the setting where only firm 1 innovates and so operates as the monopoly supplier of the product. Because $v_1 > 2$ under Assumption 1, firm 1 maximizes its profit $p_1 \min\{v_1 - p_1, 1\}$ by selling one unit to all potential customers at price $p_1 = v_1 - 1$. Firm 1's monopoly profit is thus

$$\pi_1^M = v_1 - 1\,. \tag{1}$$

Now suppose both firms innovate. Let p_i denote the price that firm $i \in \{1, 2\}$ charges for its product. Then the location of the consumer on $(0, L)$ who is indifferent between purchasing from firm 1 and from firm 2 is $l = \frac{1}{2}[\,L + v_1 - v_2 + p_2 - p_1\,]$. Therefore, the

[18] This is the equilibrium on which we focus throughout the ensuing analysis of the optimal damage rule for patent infringement. We explain below why this focus is without loss of generality, given Assumption 1.

demand functions facing firms 1 and 2 when they both serve customers and all consumers purchase one unit of a product are, respectively:

$$q_1(p_1, p_2) \;=\; \frac{1}{2}\,[\, L - \Delta + p_2 - p_1 \,] \quad \text{and} \quad q_2(p_2, p_1) \;=\; \frac{1}{2}\,[\, L + \Delta + p_1 - p_2 \,]\,. \tag{2}$$

In the absence of patent infringement, the profits (not counting innovation costs) of firms 1 and 2 are, respectively:

$$\pi_1^N \;=\; p_1\, q_1(p_1, p_2) \quad \text{and} \quad \pi_2^N \;=\; p_2\, q_2(p_2, p_1). \tag{3}$$

Firm 2's operation reduces firm 1's equilibrium profit by $R_1 = \pi_1^M - \pi_1^N$. Therefore, because firm 2 is required to pay $D = m + d_1 R_1 + d_2 \pi_2^N$ to firm 1 if firm 2 is found to have infringed firm 1's patent, the profits of firm 1 and firm 2 in the event of infringement (excluding m) are, respectively:

$$\pi_1^I \;=\; \pi_1^N + d_1 \left[\, \pi_1^M - \pi_1^N \,\right] + d_2\, \pi_2^N \quad \text{and} \quad \pi_2^I \;=\; \pi_2^N - d_1 \left[\, \pi_1^M - \pi_1^N \,\right] - d_2\, \pi_2^N. \tag{4}$$

The firms' *ex ante* expected profits (not counting innovation costs and m) when they both innovate are, respectively:

$$\pi_1^e \;=\; [\, 1 - \lambda \,]\, \pi_1^N + \lambda\, \pi_1^I \qquad \text{and} \qquad \pi_2^e \;=\; [\, 1 - \lambda \,]\, \pi_2^N + \lambda\, \pi_2^I\,. \tag{5}$$

In equilibrium, firm $i \in \{1, 2\}$ chooses p_i to maximize π_i^e, taking its rival's price as given. Letting *'s denote equilibrium outcomes, standard calculations reveal that equilibrium prices and quantities are as specified in Lemma 1.

Lemma 1. *Suppose all consumers purchase a unit of the product and each firm supplies a strictly positive level of output. Then, given d_1 and d_2, equilibrium prices and quantities are, for $i,\, j = 1, 2$ $(j \neq i)$:*

$$p_i^* \;=\; [\, 1 - \lambda\, d_j \,]\, \frac{L\,[\, 3 - \lambda\,(3\, d_i - d_j)\,] + [\, v_i - v_j \,]\,[\, 1 - \lambda\,(d_1 + d_2)\,]}{[\, 1 - \lambda\,(d_1 + d_2)\,]\,[\, 3 - \lambda\,(d_1 + d_2)\,]}\,, \quad \textit{and} \tag{6}$$

$$q_i^* \;=\; \frac{L\,[\, 3 - 2\,\lambda\, d_j \,] + v_i - v_j}{2\,[\, 3 - \lambda\,(d_1 + d_2)\,]} \;\in\; (0, L)\,, \tag{7}$$

where $\lambda\,[\, d_1 + d_2 \,] < 1$. Furthermore, $p_2^* - p_1^* \gtreqless 0$ as $\Delta \gtreqless \frac{L\,\lambda\,[\, d_2 - d_1 \,]}{2 - \lambda\,[\, d_1 + d_2 \,]}$, and all consumers *will indeed purchase a unit of the product in equilibrium if* $d_1 \geq 0$ *and* $d_2 \geq 0$ *are sufficiently small. Additionally, any prices* p_1^* *and* p_2^* *in (6) can be induced by some* d_1 *and* d_2.

Lemma 1 specifies how d_1 and d_2 affect equilibrium prices and outputs. The lemma implies that if $d_1 = d_2$, then firm i will charge a higher price than firm j in equilibrium ($p_i^* > p_j^*$) if consumers value firm i's product more highly than firm j's product (i.e., if $v_i > v_j$).

Proposition 1 examines the impact of changes in d_1 and d_2 on equilibrium outcomes, including π_i^*, the equilibrium expected profit of firm $i \in \{1, 2\}$.[19]

Proposition 1. *Following innovation by both firms: (i)* p_1^* *and* p_2^* *both increase in* d_1 *and in* d_2*; (ii)* $p_2^* - p_1^*$ *and* q_1^* *increase in* d_1*, while* $p_1^* - p_2^*$ *and* q_2^* *increase in* d_2*; (iii)* $\Pi^* \equiv \pi_1^* + \pi_2^*$ *increases in* d_1 *if* $\Delta \leq \frac{L\lambda[d_2 - d_1]}{2 - \lambda[d_1 + d_2]}$ *(so* $p_1^* \geq p_2^*$*), while* Π^* *increases in* d_2 *if* $\Delta \geq \frac{L\lambda[d_2 - d_1]}{2 - \lambda[d_1 + d_2]}$ *(so* $p_2^* \geq p_1^*$*).*

Proposition 1 reflects the following considerations. As d_1 increases, firm 2 is penalized a larger portion of firm 1's lost profit ($\pi_1^M - \pi_1^N$) if firm 2 is ultimately found to have infringed firm 1's patent. Firm 2 recognizes that it can reduce this penalty by competing less aggressively, thereby allowing firm 1's duopoly profit (π_1^N) to increase. The reduced aggression leads to higher equilibrium prices for both firms. The increased congruence of the firms' preferences regarding higher profit for firm 1 results in expanded output for firm 1, secured by a reduction in firm 1's price relative to firm 2's price.

As d_2 increases, firm 2 forfeits a larger portion of its profit (π_2^N) if it is ultimately found to have infringed firm 1's patent. Firm 1 recognizes that it can secure a larger expected penalty payment from firm 2 by competing less aggressively, thereby allowing π_2^N to increase. The reduced aggression leads to higher equilibrium prices for both firms. The increased congruence of the firms' preferences regarding higher profit for firm 2 results in expanded

[19] π_i^* is obtained by substituting p_i^* and q_i^* from equations (6) and (7) into equation (5).

output for firm 2, secured by a reduction in firm 2's price relative to firm 1's price.[20]

The increase in firm 1's equilibrium output (q_1^*) and the corresponding reduction in firm 2's output (q_2^*) induced by an increase in d_1 increase equilibrium industry profit ($\Pi^* \equiv \pi_1^* + \pi_2^*$) when firm 1's profit margin exceeds firm 2's margin (i.e., when $p_1^* > p_2^*$). Similarly, the increase in q_2^* and the reduction in q_1^* induced by an increase in d_2 increase Π^* when p_2^* exceeds p_1^*.

Before proceeding to characterize the optimal linear damage rule, it is helpful to consider the policy that maximizes welfare and the policy that maximizes industry profit following innovation by both firms. These policies are characterized in Lemmas 2 and 3, respectively. Lemma 2 refers to equilibrium welfare in the duopoly setting:

$$\widetilde{W}_{12} \;=\; v_1\, q_1^* - \int_0^{q_1^*} y\, dy + v_2\, q_2^* - \int_0^{q_2^*} y\, dy \;=\; v_1\, q_1^* - \frac{1}{2}\, q_1^{*2} + v_2\, q_2^* - \frac{1}{2}\, q_2^{*2}\,. \tag{8}$$

Lemma 2. *Suppose:*

$$d_2 \;=\; d_2^w\,(d_1) \;\equiv\; d_1 + \frac{2\,\Delta\,[\,1 - \lambda\, d_1\,]}{\lambda\,[\,L + \Delta\,]}\,. \tag{9}$$

Then, when both firms innovate, the maximum feasible level of welfare (excluding innovation costs), $v_2\, L - \frac{1}{4}\,[\,2\,L\,\Delta - \Delta^2 +\, L^2\,]$*, is induced. Equilibrium prices and quantities under (9) are:*

$$p_1^* \;=\; p_2^* \;=\; \frac{[\,1 - \lambda\, d_1\,]\,[\,L^2 - \Delta^2\,]}{L\,[\,1 - 2\,\lambda\, d_1\,] - \Delta}\,, \quad q_1^* \;=\; \frac{1}{2}\,[\,L - \Delta\,]\,, \quad \textit{and} \quad q_2^* \;=\; \frac{1}{2}\,[\,L + \Delta\,]\,. \tag{10}$$

Equilibrium industry profit is $\Pi^* = p_1^*\, L$*. This profit increases as* d_1 *or* $d_2^w(d_1)$ *increases, and reaches its maximum,* $\Pi^w = p^w\, L$*, at* $d_1 = d_1^w < \frac{1}{2\,\lambda}$*, where*

$$d_1^w \;=\; \left[\frac{1 - \Delta/L}{2\,\lambda}\right] \frac{2\,[\,v_1 - L\,] - L - \Delta}{2\,[\,v_1 - L\,] + \Delta + \frac{\Delta^2}{L}}\,, \quad \textit{so} \;\; p_1^* \;=\; p_2^* \;=\; p^w \;\equiv\; v_1 - \frac{1}{2}\,[\,L - \Delta\,]\,. \tag{11}$$

Lemma 3. *The maximum feasible level of industry profit when both firms innovate is*

$$\overline{\Pi} \;=\; \frac{L}{2}\,[\,v_1 + v_2 - L\,] + \frac{\Delta^2}{8} \;=\; \Pi^w + \frac{\Delta^2}{8}\,, \tag{12}$$

which arises when equilibrium prices are

[20] See Anton and Yao (2007), Choi (2009), and Henry and Turner (2010) for related observations.

$$\overline{p}_1 \;\equiv\; v_1 - \frac{1}{2}\left[L - \frac{\Delta}{2}\right] \quad \textit{and} \quad \overline{p}_2 \;\equiv\; v_2 - \frac{1}{2}\left[L + \frac{\Delta}{2}\right], \tag{13}$$

with corresponding equilibrium outputs $\overline{q}_1 \in (0, L)$ *and* $\overline{q}_2 = L - \overline{q}_1$*. These equilibrium prices and outputs are induced by:*

$$\overline{d}_1 \;=\; \frac{[\,L - 2\,\Delta\,]\,[\,6\,L + 7\,\Delta - 4\,v_1\,]}{4\,\lambda\,[\,2\,L^2 - 2\,L\,v_1 - 3\,\Delta^2\,]} \quad \textit{and} \quad \overline{d}_2 \;=\; \frac{[\,L + 2\,\Delta\,]\,[\,6\,L - 7\,\Delta - 4\,v_1\,]}{4\,\lambda\,[\,2\,L^2 - 2\,L\,v_1 - 3\,\Delta^2\,]}\,. \tag{14}$$

Lemmas 2 and 3 report that the welfare-maximizing duopoly prices typically differ from the prices that maximize industry profit. (They coincide only when $v_1 = v_2$.) Duopoly welfare is maximized when the two firms charge identical prices. In this event, consumers purchase from the firm that offers the largest difference between product value and transportation cost, which ensures that welfare is maximized. In contrast, industry profit is maximized when the firm with the most highly-valued product charges a higher price than its competitor. Relative to the identical welfare-maximizing prices, this price structure results in greater extraction of consumer surplus.

Lemmas 2 and 3 also report that the values of d_1 and d_2 that induce welfare-maximizing or profit-maximizing duopoly prices decline as λ increases. Thus, less stringent damages are required to secure key industry outcomes when stronger patent protection prevails.

4 Characterizing the Optimal Linear Rule

Lemma 2 identifies the welfare-maximizing allocation of industry output following innovation by both firms. To characterize the welfare-maximizing damage rule for patent infringement, one must also consider the firms' innovation incentives. To do so, observe first that welfare when only firm 1 innovates is:

$$W_1 \;=\; \widetilde{W}_1 - k_1 \;\;\text{where}\;\; \widetilde{W}_1 \;=\; v_1 - \int_0^1 y\,dy \;=\; v_1 - \frac{1}{2}\,. \tag{15}$$

Now let $\widehat{k}_i$ denote the realization of firm i's innovation cost (k_i) for which the firm will innovate if and only if $k_i \leq \widehat{k}_i$. If $\widehat{k}_1 \leq \bar{k}_1$ and $\widehat{k}_2 \leq \bar{k}_2$, then

$$\widehat{k}_1 \;=\; G(\widehat{k}_2)\,[\,\pi_1^* + \lambda\,m\,] + [\,1 - G(\widehat{k}_2)\,]\pi_1^M \quad \text{and} \quad \widehat{k}_2 \;=\; \pi_2^* - \lambda\,m\,. \tag{16}$$

Equations (8), (15), and (16) imply that *ex ante* expected welfare (i.e., the level of welfare that is anticipated before innovation costs are realized) is

$$W \;=\; F(\widehat{k}_1)\,G(\widehat{k}_2)\,\widetilde{W}_{12} + F(\widehat{k}_1)\,[\,1 - G(\widehat{k}_2)\,]\,\widetilde{W}_1 \\ -\; F(\widehat{k}_1)\int_0^{\widehat{k}_2} k_2\,dG(k_2) - \int_0^{\widehat{k}_1} k_1\,dF(k_1)\,. \tag{17}$$

To examine firm 2's innovation incentives, let k_2^w denote the increase in welfare secured by firm 2's innovation (not counting innovation costs). From equations (8) and (15):

$$k_2^w \;\equiv\; \widetilde{W}_{12} - \widetilde{W}_1 \;=\; v_1\,q_1^* - \frac{1}{2}\,q_1^{*2} + v_2\,q_2^* - \frac{1}{2}\,q_2^{*2} - (\,v_1 - \frac{1}{2}\,)\,. \tag{18}$$

Incremental welfare is maximized when firm 2 innovates if and only if its innovation cost (k_2) does not exceed k_2^w. In the ensuing analysis, k_2^{w*} will denote the increase in welfare secured by firm 2's innovation when d_2 is set (given d_1) to ensure the welfare-maximizing allocation of industry output.

To assess whether firm 1's innovation increases expected welfare, one must account for firm 2's subsequent innovation activity. Let $k_1^w(\widehat{k}_2)$ denote the increase in expected welfare secured by firm 1's innovation, given that firm 2 innovates if and only if $k_2 \leq \widehat{k}_2$. Formally:

$$k_1^w(\widehat{k}_2) \;\equiv\; G(\widehat{k}_2)\widetilde{W}_{12} + [\,1 - G(\widehat{k}_2)\,]\,\widetilde{W}_1 - \int_0^{\widehat{k}_2} k_2\,dG(k_2). \tag{19}$$

Given firm 2's innovation activity, incremental welfare is maximized when firm 1 innovates if and only if its innovation cost (k_1) does not exceed $k_1^w(\widehat{k}_2)$.

These observations imply that the optimal linear rule is the solution to problem [P], which is defined to be: $\underset{d_1,\,d_2,\,m}{Maximize}\; W$, where W is specified in equation (17), prices and quantities are specified in Lemma 1, and $\widehat{k}_i$ is specified in equation (16) for $i = 1, 2$.

The ensuing discussion will refer to the *first-best outcome* in which each firm innovates if and only if the associated increase in expected welfare exceeds the realized innovation cost, and industry output is allocated among the active producers to maximize welfare. Formally, in the first-best outcome: (i) firm 2 innovates if and only if $k_2 \;\leq\; k_2^o \;\equiv\; \min\{k_2^{w*}, \overline{k}_2\}$;

(ii) firm 1 innovates if and only if $k_1 \leq k_1^o \equiv \min\{k_1^w(k_2^o), \overline{k}_1\}$; and (from Lemma 2) (iii) $q_1^* = \frac{1}{2}[L - \Delta]$ and $q_2^* = \frac{1}{2}[L + \Delta]$ when both firms innovate.

Observation 1 considers the setting where $\widehat{k}_1 < \overline{k}_1$, so firm 1 does not always innovate because its maximum innovation cost is relatively large. The Observation reports that if the penalty for patent infringement cannot include a lump sum payment (so m is constrained to be 0) in this setting, then firm 1 innovates too seldom relative to the first-best outcome.

Observation 1. *Suppose $\widehat{k}_1 < \overline{k}_1$ and $m = 0$. Then firm 1's innovation incentive is inefficiently low (i.e., $\widehat{k}_1 < k_1^w(\widehat{k}_2)$).*

When firm 1 innovates, it creates consumer surplus that it does not fully capture, regardless of whether firm 2 innovates. Therefore, $\widetilde{W}_1 > \pi_1^M$ and $\widetilde{W}_{12} > \pi_2^* + \pi_1^*$, so aggregate welfare from firm 1's innovation always exceeds aggregate industry profit. From expression (16), when $m = 0$, firm 2 innovates if and only if $k_2 \leq \pi_2^* = \widehat{k}_2$, and firm 1 innovates if and only if $k_1 \leq \widehat{k}_1 = G(\widehat{k}_2)\,\pi_1^* + [\,1 - G(\widehat{k}_2)\,]\,\pi_1^M$. Therefore, firm 1 has inefficiently limited incentive to innovate when the penalty for patent infringement does not include a lump sum payment between the firms.

Although firm 1's innovation incentive is always inefficiently low when $m = 0$, firm 2's incentive can be excessive. If firm 2's innovation does not expand the market (so $L = 1$), then firm 2's innovation reduces firm 1's profit in the absence of patent infringement penalties. Consequently, firm 2's private benefit from innovation can exceed the corresponding social benefit (i.e., $\widehat{k}_2 > k_2^w$). In contrast, firm 2's innovation incentive can be inefficiently low (i.e., $\widehat{k}_2 < k_2^w$) if L is sufficiently large.

Observation 2. *Suppose $v_1 = v_2 \equiv v$ and $d_1 = d_2 = m = 0$. Then: (i) $k_2^w < \widehat{k}_2$ if $L = 1$; and (ii) $k_2^w > \widehat{k}_2$ if $L > 1$ and $v > \frac{1}{4}\left[\frac{3L^2 - 2}{L - 1}\right]$.*

Despite the potential conflict between the social and private incentives for innovation, the optimal linear rule can sometimes fully align these incentives while inducing the welfare-maximizing allocation of industry output. As Proposition 2 and Corollary 1 report, the optimal linear rule secures the first-best outcome if innovation costs are sufficiently small relative to the maximum feasible level of industry profit, so inequality (20) holds.

$$\Pi^w \;\geq\; \frac{k_1^o - \pi_1^M}{G(k_2^o)} + \pi_1^M + k_2^o \;\;\Leftrightarrow\;\; G(k_2^o)\,[\,\Pi^w - k_2^o\,] + [\,1 - G(k_2^o)\,]\,\pi_1^M \;\geq\; k_1^o\,. \tag{20}$$

Proposition 2. *Suppose inequality (20) holds. Then* $(\,d_1, d_2, m)$ *can be chosen to induce the first-best outcome, with* $d_2 \,=\, d_2^w(d_1)$ *and* $m \,=\, \frac{1}{\lambda}\,[\,\pi_2^* - k_2^o\,]$*. Under this (optimal) linear rule,* $d_2 \gtreqless d_1$ *as* $\Delta \gtreqless 0$*.*

Corollary 1. *If* v_1*,* v_2*, and* L *are sufficiently large to ensure* $k_2^o \,=\, \overline{k}_2$*, then the optimal linear rule secures the first-best outcome if* $\overline{k}_1 + \overline{k}_2 \;\leq\; \frac{L}{2}\,[\,v_1 + v_2 - L\,]$*. This inequality is more likely to hold if* $\overline{k}_1$ *and* $\overline{k}_2$ *are small and if* v_1*,* v_2*, and* L *are large, ceteris paribus.*

Proposition 2 and Corollary 1 consider settings where innovations are highly valued, firm 2's innovation expands the market considerably, and/or innovation costs are low. In such settings, the substantial industry profit that is potentially available can be divided between the suppliers to induce them both to always innovate even when damages are structured to induce the firms to set the same prices and thereby maximize duopoly welfare.

Proposition 2 reports that if $v_1 > v_2$, then the initial innovator's lost profit receives more weight than the second innovator's profit in the optimal linear rule (i.e., $d_1 > d_2$). In contrast, if $v_2 > v_1$, then the second innovator's profit receives more weight than the first innovator's lost profit (i.e., $d_2 > d_2$). In this sense, the optimal linear rule resembles the lost profit (LP) rule more than the unjust enrichment (UE) rule when the (vertical dimension of) quality of the initial innovation is relatively pronounced. In contrast, the optimal linear rule resembles the UE rule more than the LP rule when the quality of the follow-on innovation is relatively pronounced.

To understand the rationale for this penalty structure, recall from Proposition 1 that following innovation by both firms, the relative price of firm 1's product declines and its output increases as d_1 increases, whereas the relative price of firm 2's product declines and its output increases as d_2 increases. Therefore, the identified penalty structure helps to reduce the relative price of, and thereby shift equilibrium consumption toward, the product that consumers value most highly. Doing so ensures the welfare-maximizing allocation of industry output between the two suppliers.

Corollary 2. *If inequality (20) does not hold, then the optimal linear rule does not secure the first-best outcome.*

Corollary 2 reports that when inequality (20) does not hold, so the maximum feasible industry profit (given the welfare-maximizing output allocation) is not sufficiently large relative to innovation costs, then the optimal linear rule cannot induce welfare-maximizing innovation decisions while ensuring welfare-maximizing output allocations. The values of d_1 and d_2 required to induce the welfare-maximizing allocation of duopoly output do not generate the level of industry profit required to induce both firms to innovate whenever the social benefit of innovation exceeds the private cost of innovation.[21] In this case, the optimal linear rule will increase industry profit by eliminating the surplus of the marginal consumer. Formally, from equation (2):

$$v_1 - p_1 - \frac{1}{2}\left[\, L + v_1 - v_2 + p_2 - p_1 \,\right] \;=\; 0 \;\Rightarrow\; \; p_2 \;=\; v_1 + v_2 - p_1 - L\,. \tag{21}$$

If $v_1 \;=\; v_2 \;=\; v$, then the maximum level of industry profit that can be secured *is* $\Pi^w \;=\; \overline{\overline{\Pi}}\,$. From expression (14), the optimal linear rule sets $d_1 \;=\; d_2 \;=\; \frac{2\,v\,-\,3\,L}{4\,\lambda\,[\,v\,-\,L\,]}$ to induce the welfare-maximizing prices $p_1 \;=\; p_2 \;=\; \overline{p} \;=\; v - \frac{L}{2}$, and m is set to distribute $\overline{\overline{\Pi}}$ between the two firms to maximize W. If $v_1 \neq v_2$, then industry profit can be increased above Π^w by allowing p_1 and $p_2 \;=\; v_1 + v_2 - p_1 - L$ to diverge in order to ensure $\Pi^* \;=\; \overline{\overline{\Pi}}$.

Now consider how the optimal linear rule is structured when it cannot secure the first-best outcome. Initially, the values of p_1 and $\widehat{k}_2$ that maximize expected welfare when p_2 is

[21] When the optimal linear rule cannot secure the first-best outcome, the critical problem is to induce both firms to innovate more often. m can be adjusted to avoid settings where one firm innovates too seldom and one firm innovates too often relative to the first-best outcome. Furthermore, the logic that underlies Observation 1 explains why the critical problem is not to induce both firms to innovate less often.

set to leave the marginal consumer with zero surplus in the duopoly setting are identified (see equations (22) and (23) below), and the values of d_1 and d_2 that induce these prices are determined. m is then set to induce the identified value of $\widehat{k}_2$. (From equation (16), $m = \frac{1}{\lambda}[\,\pi_2 - \widehat{k}_2\,]$.) The properties of the optimal balanced rule then depend upon whether the identified rule generates more or less than the maximum feasible industry profit, as Proposition 3 indicates. The proposition refers to $\overline{\pi}_2$, which is firm 2's expected profit (not counting innovation costs) at the equilibrium identified in Lemma 3, where duopoly industry profit is maximized.

Proposition 3. *Suppose inequality (20) does not hold. Also suppose* $\widetilde{p}_2 = v_1 + v_2 - \widetilde{p}_1 - L$, *and* $\widetilde{p}_1$ *and* $\widetilde{k}_2 \in [\,\underline{k}_2, \overline{k}_2\,]$ *are the values of* p_1 *and* $\widehat{k}_2$ *that solve*

$$[\,k_1^w(\widehat{k}_2) - \widehat{k}_1\,]\,\frac{\partial F(\widehat{k}_1)}{\partial p_1} + F(\widehat{k}_1)\,[\,k_2^o - \widehat{k}_2\,]\,\frac{\partial G(\widehat{k}_2)}{\partial p_1} + \; F(\widehat{k}_1)\,G(\widehat{k}_2)\,\frac{\partial\,\widetilde{W}_{12}}{\partial p_1} \;=\; 0\,,\ \textit{and} \tag{22}$$

$$[\,k_1^w(\widehat{k}_2) - \widehat{k}_1\,]\,\frac{\partial F(\widehat{k}_1)}{\partial \widehat{k}_2} + F(\widehat{k}_1)\,[\,k_2^o - \widehat{k}_2\,]\,\frac{\partial G(\widehat{k}_2)}{\partial \widehat{k}_2} \;=\; 0\,. \tag{23}$$

(i) If $\widetilde{\Pi} = \widetilde{p}_1\,[\,v_1 - \widetilde{p}_1\,] + [\,v_1 + v_2 - \widetilde{p}_1 - L\,]\,[\,L + \widetilde{p}_1 - v_1\,] < \overline{\Pi}$, *then the optimal linear rule is* $(\widetilde{d}_1, \widetilde{d}_2, \widetilde{m})$, *where* $\widetilde{d}_1$ *and* $\widetilde{d}_2$ *are the values of* d_1 *and* d_2 *that induce* $\widetilde{p}_1$ *and* $\widetilde{p}_2$,[22] *and* $\widetilde{m} = \frac{1}{\lambda}[\,\widetilde{\pi}_2 - \widetilde{k}_2\,]$, *where* $\widetilde{\pi}_2 = \widetilde{p}_2\, q_2(\widetilde{p}_1, \widetilde{p}_2)$ *is firm 2's equilibrium profit.*

(ii) If $\widetilde{\Pi} \geq \overline{\Pi}$ *for the identified* $\widetilde{p}_1$ *and* $\widetilde{p}_2$, *then the optimal linear rule is* $(\overline{d}_1, \overline{d}_2, \overline{m})$, *where* $\overline{d}_1$ *and* $\overline{d}_2$ *are as specified in equation (14) (so equilibrium prices are* $\overline{p}_1$ *and* $\overline{p}_2$, *as specified in equation (13)),* $\widetilde{k}_2$ *solves equation (23), and* $\overline{m} = \frac{1}{\lambda}[\,\overline{\pi}_2 - \widetilde{k}_2\,]$.

Proposition 3 reflects the fundamental trade-off that arises when the optimal linear rule cannot secure the first-best outcome. When $v_1 \neq v_2$, setting d_1 and d_2 to induce distinct duopoly prices ($p_1 \neq p_2$) can raise industry profit, which can be employed to enhance innovation incentives. However, the distinct prices induce outputs that do not maximize welfare. Proposition 3 reports that, in effect, p_1 is optimally set to balance these two considerations

[22] These values are identified in the proof of Lemma 1 in Appendix A.

(where, given p_1, p_2 ensures that the marginal consumer receives zero surplus) as long as the resulting duopoly industry profit ($\widetilde{\Pi}$) does not exceed the maximum feasible industry profit ($\overline{\Pi}$).[23] When $\widetilde{\Pi} \geq \overline{\Pi}$, profit cannot be increased further by altering equilibrium prices, so the maximum industry profit constraint binds. The optimal prices maximize industry profit in this case (i.e., $p_1 = \overline{p}_1$ and $p_2 = \overline{p}_2$).[24]

$k_1^w - \widehat{k}_1 > 0$ can be viewed as a measure of the extent to which firm 1 has insufficient incentive to innovate. $F(\widehat{k}_1)\,[\,k_2^w - \widehat{k}_2\,]$ can be viewed as a measure of the extent to which firm 2 is expected to have insufficient incentive to innovate, where the expectation reflects the probability that firm 2 will have an opportunity to innovate (because firm 1 has innovated). Equation (23) implies that the patent infringement penalty is optimally increased to the point where the marginal reduction in firm 1's innovation deficiency is equal to the increase in firm 2's expected innovation deficiency, taking into account the rate at which $\widehat{k}_1$ declines and $\widehat{k}_2$ increases as the penalty increases.

5 The Optimality of the Linear Rule

We now demonstrate that the optimal linear rule achieves the highest expected welfare among all balanced patent infringement damage rules. A balanced damage rule can specify the prices the firms set, a transfer payment from firm 2 to firm 1 following innovation by both firms, the probability that firm 1 innovates, and the probability that firm 2 innovates, following innovation by firm 1.[25] By the revelation principle (e.g., Myerson, 1979), it is without loss of generality to consider truthful direct mechanisms, where firms are induced to truthfully report their privately-observed innovation costs.

[23] Notice that if $v_1 = v_2$, then $\widetilde{p}_1 = \widetilde{p}_2 = p^w$ and $\widetilde{\Pi} = \overline{\Pi} = \Pi^W$, so only equation (23) is relevant.

[24] Assumption 1 ensures that all consumers purchase a unit of the product and each firm serves some customers in both the welfare-maximizing and profit-maximizing outcomes. Consequently, Proposition 3 implies that when characterizing the optimal linear rule, there is no loss of generality in focusing on the duopoly equilibrium in which all consumers purchase a unit of the product and both firms serve some customers.

[25] The balanced damage rules that we analyze permit the specification of prices, not quantities. However, the specification of a sufficiently high price for firm 2's product can effectively preclude the sale of the product, thereby functioning like an injunction (e.g., Shapiro, 2016).

Let $\alpha_1(k_1') \in [0,1]$ denote the probability that firm 1 is required to innovate when it reports its innovation cost to be k_1'. Also let $\alpha_2(k_1', k_2') \in [0,1]$ denote the probability that firm 2 is required to innovate following innovation by firm 1 when firm i reports its innovation cost to be k_i', for $i = 1, 2$. $p_i(k_1', k_2')$ will denote the corresponding price that firm $i \in \{1, 2\}$ must set and $T(k_2', k_2')$ will denote the corresponding payment that firm 2 must make to firm 1 when both firms innovate. $p_1^M(k_1', k_2')$ will denote the corresponding price that firm 1 must set when it is the sole innovator. Denote this vector of policy instruments by:

$$Z(k_1', k_2') \equiv \left(\alpha_1(k_1'),\, \alpha_2(k_1', k_2'),\, p_1(k_1', k_2'),\, p_2(k_1', k_2'),\, p_1^M(k_1', k_2'),\, T(k_1', k_2') \right).$$

In addition, let $q_i^z(k_1', k_2')$ denote the demand for firm i's product following innovation by both firms, given that firm $i \in \{1, 2\}$ sets price $p_i(k_1', k_2')$. Furthermore, let $\pi_i^z(k_1', k_2') = p_i(k_1', k_2')\, q_i^z(k_1', k_2')$ denote the corresponding profit of firm i (not counting innovation costs). $q_1^{Mz}(k_1', k_2')$ will denote the demand for firm 1's product when it is the monopoly supplier and it sets price $p_1^M(k_1', k_2')$. $\pi_1^{Mz}(k_1', k_2') = p_1^M(k_1', k_2')\, q_1^{Mz}(k_1', k_2')$ will denote firm 1's corresponding monopoly profit (not counting innovation costs).

Total welfare in this setting when innovation costs (k_1, k_2) are reported truthfully is:

$$\begin{aligned} W^Z &= \int_0^{\overline{k}_1} \alpha_1(k_1) \{ \int_0^{\overline{k}_2} \{ \alpha_2(k_1, k_2) [\widetilde{W}_{12}^z(k_1, k_2) - k_2] \\ &\qquad + [1 - \alpha_2(k_1, k_2)] \widetilde{W}_1^z(k_1, k_2) \} \, dG(k_2) - k_1 \} \, dF(k_1) \end{aligned}$$

$$\text{where} \quad \widetilde{W}_1^z(k_1, k_2) = v_1\, q_1^{Mz}(k_1, k_2) - \frac{1}{2} [q_1^{Mz}(k_1, k_2)]^2 \quad \text{and}$$

$$\widetilde{W}_{12}^z(k_1, k_2) = v_1\, q_1^z(k_1, k_2) - \frac{1}{2} [q_1^z(k_1, k_2)]^2 + v_2\, q_2^z(k_1, k_2) - \frac{1}{2} [q_2^z(k_1, k_2)]^2 .$$

Firm 1's expected payoff when its innovation cost is k_1, it reports this cost to be k_1', and it anticipates that firm 2 will report its innovation cost truthfully is:

$$B_1(k_1' \mid k_1) \equiv \alpha_1(k_1') [H(k_1') - k_1], \text{ where}$$

$$\begin{aligned} H(k_1') &\equiv \int_0^{\overline{k}_2} \{ \alpha_2(k_1', k_2) [\pi_1^z(k_1', k_2) + T(k_1', k_2)] \\ &\qquad + [1 - \alpha_2(k_1', k_2)] \pi_1^{Mz}(k_1', k_2) \} \, dG(k_2) . \end{aligned} \tag{24}$$

Following report k_1' by firm 1, firm 2's expected payoff when its innovation cost is k_2 and it reports this cost to be k_2' is:

$$B_2(k_2' \,|\, k_2; k_1') \;\equiv\; \alpha_2(k_1', k_2')\,[\,\pi_2^z(k_1', k_2') - T(k_1', k_2') - k_2\,]\,. \tag{25}$$

The welfare-maximizing damage policy in this setting is the solution to the following problem, denoted [P-Z]:

$$\underset{Z(k_1,k_2)}{Maximize} \quad W^Z, \quad \text{subject to, for all } k_i,\;\; k_i' \in [\,0, \bar{k}_i\,]\ (i \in \{1,2\}):$$

$$B_1(k_1 \,|\, k_1) \;\geq\; \text{maximum}\,\{\,0\,, B_1(k_1' \,|\, k_1)\,\}\,, \;\text{ and} \tag{26}$$

$$B_2(k_2 \,|\, k_2; k_1) \;\geq\; \text{maximum}\,\{\,0\,, B_2(k_2' \,|\, k_2; k_1)\,\}\,. \tag{27}$$

Inequality (26) ensures that firm 1 truthfully reports its realized innovation cost and secures a nonnegative expected payoff by doing so. Inequality (27) ensures the corresponding outcomes for firm 2, for any cost report by firm 1.

To characterize the solution to [P-Z], it is helpful to consider problem [P-Z]$'$, which is [P-Z] except that constraints (26) and (27) are replaced by:

$$B_1(k_1 \,|\, k_1) \;\geq\; 0 \quad \text{and} \quad B_2(k_2 \,|\, k_2; k_1) \;\geq\; 0\,. \tag{28}$$

Observe that if constraints (26) and (27) are satisfied at a solution to [P-Z]$'$, then the identified solution to [P-Z]$'$ is a solution to [P-Z].

Lemma 4. *$p_1^M(k_1, k_2) = p_1^M = v - 1$ for all $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$ for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) = 0$ at a solution to [P-Z]$'$. Moreover, $q_i^z(k_1, k_2) = q_i(p_1(k_1, k_2), p_2(k_1, k_2))$ as specified in equation (2) for $i = 1, 2$, for all $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$ for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$ at a solution to [P-Z]$'$.*

Lemma 4, which reflects Assumption 1, indicates that full market coverage is induced in both the monopoly and duopoly settings at the solution to [P-Z]$'$.

Lemma 5. *At a solution to [P-Z]$'$, there exist $\tilde{k}_1 \in [\,0, \overline{k}_1\,]$ and $\tilde{k}_2 \in [\,0, \overline{k}_2\,]$ such that: (i) $\alpha_1(k_1) = 1$ for all $k_1 \in [\,0, \tilde{k}_1\,]$ and $\alpha_1(k_1) = 0$ for all $k_1 \in (\tilde{k}_1, \overline{k}_1\,]$; (ii) for each*

$k_1 \in [0, \tilde{k}_1]$: $\alpha_2(k_1, k_2) = 1$ for all $k_2 \in [0, \tilde{k}_2]$ and $\alpha_2(k_1, k_2) = 0$ for all $k_2 \in (\tilde{k}_2, \overline{k}_2]$; and (iii) $p_i(k_1', k_2') = p_i(k_1'', k_2'')$ and $T(k_1', k_2') = T(k_1', k_2'')$ for all k_2', $k_2'' \in [0, \tilde{k}_2]$ for each $k_1' \in [0, \tilde{k}_1]$.

Lemma 5 indicates that if innovation is induced, it is induced for the smaller realizations of the firms' innovation costs. Furthermore, stochastic innovation ($\alpha_i(\cdot) \in (0, 1)$) serves no useful purpose in the present setting. In addition, because the firms' production costs do not vary with their innovation costs, there is no gain from inducing duopoly prices and output levels (and thus transfer payments, $T(\cdot)$) that vary with reported innovation costs.

Lemmas 4 and 5 imply that each firm effectively faces the simple choice of innovating or not innovating at the identified solution to [P-Z]$'$. The choices are structured so that innovation is profitable for a firm if and only if its realized innovation cost is sufficiently small, i.e., if $k_i \leq \tilde{k}_i$. Therefore, the firms have no incentive to misrepresent their realized innovation costs, so constraints (26) and (27) are satisfied at the identified solution to [P-Z]$'$. Consequently, a solution to [P-Z]$'$ that satisfies the properties identified in Lemmas 4 and 5 is a solution to [P-Z].

It follows that expected welfare in the present setting can be written as:

$$\widetilde{W}^Z = F(\tilde{k}_1)\, G(\tilde{k}_2)\, \widetilde{W}_{12} + F(\tilde{k}_1)\, [\, 1 - G(\tilde{k}_2)\,]\, \widetilde{W}_1 - F(\tilde{k}_1) \int_0^{\tilde{k}_2} k_2 \, dG(k_2) - \int_0^{\tilde{k}_1} k_1 \, dF(k_1)\,. \tag{29}$$

Furthermore, when $\tilde{k}_i \leq \overline{k}_i$ for $i = 1, 2$:

$$[\, \pi_1^* + T\,]\, G(\tilde{k}_2) + \pi_1^M [\, 1 - G(\tilde{k}_2)\,] = \tilde{k}_1 \quad \text{and} \quad \pi_2^* - T = \tilde{k}_2\,, \tag{30}$$

where T is a constant and $\pi_i^* = p_i\, q_i(p_1, p_2)$ does not vary with k_1 or k_2, for $i \in \{1, 2\}$.

Because the corresponding industry profit is $\Pi^* = \pi_1^* + T + (\pi_2^* - T) = \pi_1^* + T + \tilde{k}_2$, expression (30) can be written as:

$$[\, \Pi^* - \tilde{k}_2\,]\, G(\tilde{k}_2) + \pi_1^M\, [\, 1 - G(\tilde{k}_2)\,] = \tilde{k}_1 \quad \text{and} \quad \pi_2^* - T = \tilde{k}_2\,. \tag{31}$$

Therefore, problem [P-Z]$'$ can be written as problem [P-Z]$''$: $\underset{p_1, p_2, T}{Maximize}\ \widetilde{W}^Z$, where $\widetilde{W}^Z$ is specified in equation (29), the firms' outputs are specified in expression (2), and $\tilde{k}_i$ is specified in expression (31) when $\tilde{k}_i \leq \overline{k}_i$, for $i = 1, 2$. Consequently, problem [P-Z]$''$ is identical to Problem [P], which ensures the following conclusion holds.

Proposition 4. *The optimal linear rule achieves the highest welfare among all balanced damage policies.*

Proposition 4 reflects the fact that linear rules can link damage payments to the profits of both firms and can specify lump-sum transfer payments between the firms. Consequently, linear rules provide widespread latitude to induce desired allocations of industry output while implementing desired configurations of industry profit and corresponding innovation incentives.

Before concluding, we illustrate the welfare gains that the optimal linear rule can secure relative to the LP rule and the UE rule. We do so in the following *baseline setting*, where consumers value the two firms' products symmetrically, the innovation costs for the two firms have the same uniform distribution, and firm 2's innovation increases the size of the market by eighty percent.

Baseline Setting. $v_1 = v_2 = 7.5$, $L = 1.8$, $\lambda = 0.50$, and $F(k_1)$ and $G(k_2)$ are uniform distributions with $\underline{k}_i = 0$ and $\overline{k}_i = 5$, for $i = 1, 2$.

Table 1 identifies the optimal linear rule for patent infringement (d_1^*, d_2^*, m^*) in the baseline setting and as product valuations change.[26] The table also reports the level of expected welfare that arises under the optimal linear rule (W^*), under the UE rule (W^{UE}), under the LP rule (W^{LP}), and in the first-best outcome (W^{FB}).

[26] As the product valuations, v_1 and v_2, change in Table 1, the values of all other parameters remain at their values in the baseline setting.

v_1	v_2	d_1^*	d_2^*	m^*	W^*	W^{UE}	W^{LP}	W^{FB}
5	5	0.72	0.72	0.21	3.03	2.49	2.58	3.36
7	8	0.25	1.5	5.56	7.94	6.96	5.35	7.94
7.5	7.5	0.81	0.81	0	7.69	5.98	6.22	7.69
8	7	1.5	0.25	-5.56	7.94	6.37	6.16	7.94
10	10	0.81	0.81	0	12.19	10.03	9.45	12.19

Table 1. The Effects of Changing Product Valuations.

Four elements of Table 1 warrant emphasis. First, there are many settings where the optimal linear rule ensures the first-best outcome. Second, the optimal linear rule often secures a substantial increase in welfare above the levels generated by the LP and UE rules. This is the case both when the optimal linear rule achieves the first-best outcome and when it does not do so. To illustrate, W^* exceeds $\max\{W^{UE}, W^{LP}\}$ by more than 21% when $v_1 = v_2 = 10$, and by more than 17% when $v_1 = v_2 = 5$. Also observe that W^* exceeds W^{LP} by more than 48% when $v_1 = 7$ and $v_2 = 8$.

Third, the lump-sum payment from firm 2 to firm 1 under the optimal linear rule (m^*) can be positive, negative, or zero. For the settings in Table 1 where $v_1 + v_2 = 15$, m^* is positive (negative) when consumers value firm 2's (firm 1's) product relatively highly and is zero when $v_1 = v_2$.[27]

Fourth, the optimal linear rule does not always resemble the UE rule more than the LP rule (in the sense that $d_2^* > d_1^*$) when the UE rule generates a higher level of welfare than the LP rule.[28] This is the case because d_1^* and d_2^* affect multiple determinants of welfare – both output allocations and innovation decisions – in nonlinear fashion. Consequently, even though welfare might be higher when, say, $d_2 = 1$ and $d_1 = m = 0$ than when $d_1 = 1$ and $d_2 = m = 0$, it does not follow that d_2 will exceed d_1 under the optimal linear rule.

[27] Observe that when $v_1 + v_2 = 15$, W^* is higher in Table 1 when $v_1 \neq v_2$ than when $v_1 = v_2$. Welfare is higher in the presence of asymmetric product valuations here because consumers purchase more of the product they value most highly and less of the product they value less highly.

[28] See the fourth row of data in Table 1 where $v_1 = 8$ and $v_2 = 7$.

Appendix B illustrates how the optimal linear rule and its performance vary as other model parameters ($\overline{k}_1, \overline{k}_2, L$, and λ) change. The numerical solutions reported in Appendix B indicate, for example, that the relative performance of the optimal linear rule (W^*/W^{LP} and W^*/W^{UE}) tends to increase as L (the extent to which firm 2's innovation expands the market) increases.

6 Conclusion

We have introduced and characterized the optimal linear rule for patent infringement damages under sequential innovation. We have demonstrated that this rule often secures substantially higher welfare than common rules like the LP rule and the UE rule. We have also identified conditions under which a linear rule can induce both efficient incentives for sequential innovation and the efficient allocation of industry output. Moreover, we have shown that the linear rule is optimal among all balanced patent infringement damage rules.

We have emphasized the welfare gains that the optimal linear rule can secure relative to the LP rule and the UE rule. However, the optimal linear rule can also secure substantially higher levels of welfare than other popular damage rules, including the *reasonable royalty* (RR) rule. Under the RR rule, when firm 2 is found to have infringed firm 1's patent, firm 2 is required to deliver to firm 1 the royalty payments that firm 1 would have collected if the two firms had negotiated a licensing/royalty agreement, knowing that firm 2's product infringes firm 1's patent (e.g., Henry and Turner, 2010).[29] Absent contracting frictions, the firms would negotiate an agreement that maximizes industry profit in this setting. Therefore, under the RR rule, duopoly industry profit will be $\overline{\Pi}$. Furthermore, following firm 1's innovation, firm 2 will innovate if and only if the follow-on innovation increases industry profit (so $k_2 \leq \overline{\Pi} - \pi_1^M \equiv k_2^R$). Industry profit can differ from $\overline{\Pi}$ under the optimal linear rule, and k_2^R often differs from the values of k_2^o and $\tilde{k}_2$ specified in Proposition 3. Therefore,

[29] As noted above, U.S. patent law stipulates that the damage penalty for patent infringement must be "adequate to compensate for the infringement, but in no event less than a reasonable royalty for the use made of the invention by the infringer, together with interest and costs as fixed by the court" (Frank and DeFranco, 2000-2001, p. 281).

the optimal linear rule often secures a strictly higher level of welfare than the RR rule (and always secures at least as high a level of welfare).

The expanded Hotelling model that we analyzed facilitates the characterization of the optimal linear rule. The optimality of the linear rule and the substantial welfare gains it can secure relative to the LP, UE, and RR rules seem likely to persist more generally. To illustrate, suppose the duopoly demand for the differentiated product of firm $i \in \{1, 2\}$ is $q_i(p_1, p_2; L)$, where L is a parameter that measures the extent to which firm 2's innovation expands the market (so $\frac{\partial q_i(\cdot)}{\partial L} > 0$). In this setting, the values of d_1 and d_2 in the optimal linear rule will continue to induce each firm to partially internalize its rival's profit. Therefore, higher values of d_1 and d_2 will continue to induce higher equilibrium prices and profit, and m can be employed to allocate realized profit to secure desired innovation incentives. One difference that will arise in this more general setting is that deadweight loss typically will arise when prices diverge from marginal cost. Consequently, although the optimal linear rule is likely to continue to outperform other damage rules (including the LP, UE, and RR rules), it typically will not secure the first-best outcome, as it often does in the expanded Hotelling model.

For simplicity, we have considered settings where key parameters are assumed to be known to all parties from the outset of their interaction. More generally, uncertainty about key parameters (e.g., v_1, v_2, and L) or the ultimate success of innovation activity may prevail when innovation decisions are made. In such settings, the optimal linear rule may not ensure that welfare is maximized *ex post*. However, it seems likely to continue to secure higher levels of expected welfare than popular damage rules in light of its expanded capacity to influence aggregate industry profit, and to allocate both industry profit and industry output between sequential innovators.

The probability of patent infringement, λ, was taken to be an exogenous parameter in our model. In practice, this probability can vary with L, v_1, and v_2. The foregoing analysis could readily incorporate this dependence, since our key findings hold for any specified value of λ. More generally, λ might be an element of a patent protection policy that is chosen optimally, along with d_1, d_2, and m.[30] As noted at the outset, the economics literature has devoted considerable attention to the design of patent protection policy. We have shown that for any level of patent protection $\lambda > 0$, the optimal linear rule ensures the first-best outcome when innovation costs are not too large. In contrast, common damage rules (e.g., the LP, UE, and RR rules) typically cannot induce the ideal outcome for any level of patent protection. These observations suggest the importance of devoting additional attention to patent infringement damage rules in the optimal design and enforcement of patent policy.

Future research might extend our analysis to allow for more than two firms, more than two rounds of innovation, and limited innovator wealth. These extensions may limit the circumstances under which efficient innovation decisions and efficient output allocations can be ensured simultaneously.[31] However, the extensions seem unlikely to alter our findings regarding the superiority of an optimally designed linear rule, given its ability to secure desired levels of industry profit, distribute the profit to foster innovation by multiple firms, and ensure welfare-maximizing allocations of industry output.

[30] A complete analysis of the optimal value of λ would need to consider, among other things, the social costs of detecting patent infringement and the costs associated with firms' efforts to "invent around" patents (e.g., Gallini, 1992).

[31] For instance, the welfare-maximizing values of m and d_1 may not be feasible when firm 2's financial resources are limited. In this case, d_2 may optimally be increased to help offset the reductions in m and d_1. In settings where λ is a policy instrument, relatively large values of λ may optimally accompany the small values of m and d_1 dictated by firm 2's limited financial resources.

Appendix A

This Appendix presents the proofs of the formal conclusions in the text.

Proof of Lemma 1

From equations (2) – (5), the expected profit of firm 1 and firm 2 are, respectively

$$\pi_1^e \;=\; [\,1-\lambda\,d_1\,]\,\frac{p_1}{2}\,[\,L-\Delta+p_2-p_1\,]+\lambda\left[d_1\,\pi_1^M+d_2\,p_2\,\frac{1}{2}\,(L+\Delta+p_1-p_2)\right],$$

$$\pi_2^e \;=\; [\,1-\lambda\,d_2\,]\,\frac{p_2}{2}\,[\,L+\Delta+p_1-p_2\,]-\lambda\,d_1\left[\pi_1^M-p_1\,\frac{1}{2}\,(L-\Delta+p_2-p_1)\right].$$

In an equilibrium where all consumers purchase a unit of the product and both firms sell positive levels of output, the first-order conditions for p_1 and p_2 are:

$$\frac{\partial \pi_1^e}{\partial p_1} \;=\; \frac{1}{2}\,[\,1-\lambda\,d_1\,]\,[\,p_2-2\,p_1+v_1-v_2+L\,]+\frac{\lambda}{2}\,d_2\,p_2 \;=\; 0\,, \tag{32}$$

$$\frac{\partial \pi_2^e}{\partial p_2} \;=\; \frac{1}{2}\,[\,1-\lambda\,d_2\,]\,[\,p_1-2\,p_2+v_2-v_1+L\,]+\frac{\lambda}{2}\,d_1\,p_1 \;=\; 0\,. \tag{33}$$

The corresponding second-order conditions are $\lambda\,d_1 < 1$ and $\lambda\,d_2 < 1$. Solving (32) and (33) provides the equilibrium prices p_1^* and p_2^* specified in (6). The equilibrium outputs q_1^* and q_2^* in (7) then follow from (2).

To ensure p_1^* and p_2^* are non-negative and finite when all consumers purchase a unit of the product, it must be the case that $\lambda\,[\,d_1+d_2\,] < 1$. Suppose, to the contrary, that $\lambda\,[\,d_1+d_2\,] \geq 1$. Observe that $3-\lambda\,[\,d_1+d_2\,] > 0$ because $\lambda\,d_1 < 1$ and $\lambda\,d_2 < 1$. Therefore, when $\lambda\,[\,d_1+d_2\,] \geq 1$, the denominator in (6) is

$$[\,1-\lambda\,(d_1+d_2)\,]\,[\,3-\lambda\,(d_1+d_2)\,] \;\leq\; 0\,. \tag{34}$$

If $\Delta \geq 0$, then

$$L\,[\,3-\lambda\,(3\,d_1-d_2)\,]-\Delta\,[\,1-\lambda\,(d_1+d_2)\,] \;\geq\; L\,[\,3-\lambda\,(3\,d_1-d_2)\,] \;>\; 0\,.$$

Consequently, (6) and (34) imply that $p_1^* < 0$ or $p_1^* = \infty$, which is a contradiction. If $\Delta < 0$, then

$$L\,[\,3-\lambda\,(3\,d_2-d_1)\,]+\Delta\,[\,1-\lambda\,(d_1+d_2)\,] \;\geq\; L\,[\,3-\lambda\,(3\,d_2-d_1)\,] \;>\; 0\,.$$

Therefore, (6) and (34) imply that $p_2^* < 0$ or $p_2^* = \infty$, which is a contradiction.

From (7):

$$q_1^* \;=\; \frac{3\,L-\Delta-2\,L\,\lambda\,d_2}{2\,[\,3-\lambda\,(d_1+d_2)\,]}\,.$$

$q_1^* > 0$ because $3\,L-\Delta-2\,L\,\lambda\,d_2 \;=\; L\,[\,3-2\,\lambda\,d_2\,]-\Delta \;>\; 0$, since $\lambda\,d_2 < 1$ and $\Delta < L$ from Assumption 1. $q_1^* < L$ because

$$3\,L-\Delta-2\,L\,\lambda\,d_2-2\,[\,3-\lambda\,(d_1+d_2)\,]\,L \;=\; -\,L\,[\,3-2\,\lambda\,d_2\,]+\Delta \;<\; 0\,.$$

Therefore, $q_1^* \in (0, L)$, and so $q_2^* = L - q_1^* \in (0, L)$.

Furthermore, from (6),

$$\begin{aligned} p_2^* - p_1^* &= \frac{L\lambda\,[\,d_2 - d_1\,]\,[\,\lambda\,(\,d_1 + d_2\,) - 1\,] + \Delta\,\{\,2 + \lambda\,(d_1 + d_2)\,[\,\lambda\,(\,d_1 + d_2\,) - 3\,]\,\}}{[\,1 - \lambda\,(d_1 + d_2)\,]\,[\,3 - \lambda\,(d_1 + d_2)\,]}. \\ &\stackrel{s}{=} L\lambda\,(d_2 - d_1)\,[\,\lambda\,(\,d_1 + d_2\,) - 1\,] + \Delta\,\{\,2 + \lambda\,(d_1 + d_2)\,[\,\lambda\,(\,d_1 + d_2\,) - 3\,]\,\} \\ &= [\,1 - \lambda\,(d_1 + d_2)\,]\,[\,L\,\lambda\,(d_1 - d_2) + \Delta\,(\,2 - \lambda\,[\,d_1 + d_2\,]\,)\,] \\ &\gtreqless 0 \quad \Leftrightarrow \quad \Delta \gtreqless \frac{L\,\lambda\,[\,d_2 - d_1\,]}{2 - \lambda\,[\,d_1 + d_2\,]}\,. \end{aligned} \tag{35}$$

Observe from (6) that when $d_1 = d_2 = 0$

$$v_2 - p_2 - q_2^* = v_2 - 2\,L + \frac{1}{2}\,[\,L - \Delta\,] > 0$$

because $\min\{v_1, v_2\} > 2\,L$ and $L - \Delta > 0$, from Assumption 1. By continuity, if $d_1 \geq 0$ and $d_2 \geq 0$ are sufficiently small, then consumers will continue to secure strictly positive surplus from purchasing a unit of the product.

Finally, consider

$$d_1\,(p_1, p_2) = \frac{[\,L + \Delta + p_1 - 2\,p_2\,]\,[\,L - \Delta - 2\,p_1 + 2\,p_2\,]}{\lambda\,[\,L^2 - L\,p_1 - L\,p_2 + 3\,\Delta\,p_2 - 3\,\Delta\,p_1 + 4\,p_1\,p_2 - 2\,p_1^2 - 2\,p_2^2 - \Delta^2\,]}, \text{ and}$$

$$d_2\,(p_1, p_2) = \frac{[\,L + \Delta + 2\,p_1 - 2\,p_2\,]\,[\,L - \Delta - 2\,p_1 + p_2\,]}{\lambda\,[\,L^2 - L\,p_1 - L\,p_2 + 3\,\Delta\,p_2 - 3\,\Delta\,p_1 + 4\,p_1\,p_2 - 2\,p_1^2 - 2\,p_2^2 - \Delta^2\,]}\,. \tag{36}$$

It can be verified that if these values of d_1 and d_2 are substituted into the expressions for p_1 and p_2 in (6), the resulting expressions are indeed p_1 and p_2. Therefore, these values of d_1 and d_2 induce the equilibrium prices specified in (6). ■

Proof of Proposition 1

From (7):

$$\frac{\partial q_1^*}{\partial d_1} = \frac{\lambda\,q_1^*}{3 - \lambda\,d_1 - \lambda\,d_2} > 0 \quad \text{and} \quad \frac{\partial q_1^*}{\partial d_2} = \frac{-\,\lambda\,q_2^*}{3 - \lambda\,d_1 - \lambda\,d_2} < 0\,. \tag{37}$$

The inequalities in (37) hold because $\lambda\,[\,d_1 + d_2\,] < 1$ from Lemma 1. (37) implies that $\frac{\partial q_2^*}{\partial d_1} < 0$ and $\frac{\partial q_2^*}{\partial d_2} > 0$ because $q_2^* = L - q_1^*$. From (2):

$$\frac{\partial q_1^*}{\partial d_i} = \frac{1}{2}\,\frac{\partial\,(p_2^* - p_1^*)}{\partial d_i}\,; \qquad \frac{\partial q_2^*}{\partial d_i} = \frac{1}{2}\,\frac{\partial\,(p_1^* - p_2^*)}{\partial d_i} \quad \text{for } i = 1, 2\,. \tag{38}$$

Therefore $\frac{\partial\left(p_2^* - p_1^*\right)}{\partial d_i} \stackrel{s}{=} \frac{\partial q_1^*}{\partial d_i}$ and $\frac{\partial\left(p_1^* - p_2^*\right)}{\partial d_i} \stackrel{s}{=} \frac{\partial q_2^*}{\partial d_i}$.

Consider the case where $d_1 \geq d_2$. Observe that:

$$\lambda\,[\,2 - \lambda\,(d_1 + d_2)\,]\,\min\,\{\,5\,d_2 - 3\,d_1,\ 5\,d_1 - 3\,d_2\} + 3$$

$$\geq\ \lambda\,[\,2 - \lambda\,(d_1 + d_2)\,]\,[\,-3\,d_1] + 3\ =\ 3\,[\,1 - \lambda\,d_1\,(\,2 - \lambda\,[\,d_1 + d_2\,]\,)\,]\,. \tag{39}$$

$\frac{|\,\Delta\,|}{L} \leq 1$, from Assumption 1. Therefore, (39) implies:

$$\lambda\,[\,2 - \lambda\,(d_1 + d_2)\,]\,\min\,\{\,5\,d_2 - 3\,d_1,\ 5\,d_1 - 3\,d_2\,\} + 3\ >\ \frac{|\,\Delta\,|}{L}\,[\,1 - \lambda\,(d_1 + d_2)\,]^2 \tag{40}$$

$$\text{if}\quad 3\,[\,1 - \lambda\,d_1\,(2 - \lambda\,[\,d_1 + d_2\,])\,]\ >\ [\,1 - \lambda\,(d_1 + d_2)\,]^2$$

$$\Leftrightarrow\ -\lambda^2 d_2^2 + 2\,\lambda\,d_2 + \lambda^2\,d_1\,d_2 + 2\left[\,1 - 2\,\lambda\,d_1 + \lambda^2 d_1^2\,\right]\ >\ 0\,. \tag{41}$$

The inequality in (41) holds because $d_1 \geq d_2$ by assumption and $1 - 2\,\lambda\,d_1 + \lambda^2 d_1^2 = [\,1 - \lambda\,d_1\,]^2 > 0\,.$

From (6) and (40):

$$\frac{\partial p_1^*}{\partial d_1}\ =\ [\,1 - \lambda\,d_2\,]\,\lambda\,\frac{L\,(\,[\,2 - \lambda\,(d_1 + d_2)]\,[\,5\,d_2 - 3\,d_1\,] + 3\,) - \Delta\,[\,1 - \lambda\,(d_1 + d_2)\,]^2}{[\,1 - \lambda\,(d_1 + d_2)\,]^2\,[\,3 - \lambda\,(d_1 + d_2)\,]^2}\ >\ 0\,;$$

$$\frac{\partial p_2^*}{\partial d_2}\ =\ [\,1 - \lambda\,d_1\,]\,\lambda\,\frac{L\,(\,[\,2 - \lambda\,(d_1 + d_2)\,]\,[\,5\,d_1 - 3\,d_2\,] + 3) + \Delta\,[\,1 - \lambda\,(d_1 + d_2)\,]^2}{[\,1 - \lambda\,(d_1 + d_2)\,]^2\,[\,3 - \lambda\,(d_1 + d_2)\,]^2}\ >\ 0\,.$$

Moreover, $\frac{\partial p_2^*}{\partial d_1} > 0$ and $\frac{\partial p_1^*}{\partial d_2} > 0$ because $\frac{\partial\left(p_2^* - p_1^*\right)}{\partial d_1} > 0$ and $\frac{\partial\left(p_1^* - p_2^*\right)}{\partial d_2} > 0\,.$

From Lemma 1:

$$\Pi^*\ =\ \pi_1^* + \pi_2^*\ =\ p_1^*\,q_1^* + p_2^*\,[\,L - q_1^*\,]\ =\ [\,p_1^* - p_2^*\,]\,q_1^* + p_2^*\,L\,. \tag{42}$$

Because $\frac{\partial q_2^*}{\partial d_1} < 0$, (35) and (42) imply that if $\Delta\,[\,2 - \lambda\,(\,d_1 + d_2\,)\,] \leq L\,\lambda\,[\,d_2 - d_1\,]$, then:

$$\frac{\partial \Pi^*}{\partial d_1}\ =\ \frac{\partial\,(p_2^* - p_1^*)}{\partial d_1}\,q_2^* + [\,p_2^* - p_1^*\,]\,\frac{\partial q_2^*}{\partial d_1} + \frac{\partial p_1^*}{\partial d_1}\,L\ >\ 0\,.$$

Finally, if $\Delta\,[\,2 - \lambda\,(d_1 + d_2)\,] \geq L\,\lambda\,[\,d_2 - d_1\,]$, then $p_2^* - p_1^* \geq 0$ from (35) and $\frac{\partial q_1^*}{\partial d_2} < 0$ from (37). Therefore:

$$\frac{\partial \Pi^*}{\partial d_2}\ =\ \frac{\partial\,(p_1^* - p_2^*)}{\partial d_2}\,q_1^* + [\,p_1^* - p_2^*\,]\,\frac{\partial q_1^*}{\partial d_2} + \frac{\partial p_2^*}{\partial d_2}\,L\ >\ 0\,.$$

The proof for the case where $d_1 < d_2$ is analogous. ■

Proof of Lemma 2

Substituting d_1 and $d_2^w(d_1)$ into (6) and (7) provides:

$$p_1^* = p_2^* = \frac{[\,1-\lambda\, d_1\,]\,[\,L^2-\Delta^2\,]}{L\,[\,1-2\,\lambda\, d_1\,]-\Delta}\,,\ \ q_1^* = \frac{1}{2}\,[\,L-\Delta\,]\,,\ \text{ and }\ q_2^* = \frac{1}{2}\,[\,L+\Delta\,]\,.$$

Equilibrium industry profit is

$$\Pi^* = p_1^*\, q_1^* + p_2^*\, q_2^* = p_1^*\, L = \frac{[\,1-\lambda\, d_1\,]\,[\,L^2-\Delta^2\,]}{1-2\,\lambda\, d_1-\Delta/L}\,,$$

which increases as d_1 (or d_2^w) increases.

The highest industry profit is attained when the marginal consumer has zero surplus which, from (10), occurs when

$$v_2 - p_2^* - q_2^* = v_2 - p_2^* - \frac{L+\Delta}{2} = 0 \ \Leftrightarrow\ p_2^* = p_1^* = v_2 - \frac{1}{2}\,[\,L+\Delta\,] \equiv p^w\,. \quad (43)$$

(43) implies that equilibrium industry profit in this case is $\Pi^w = p^w\, L$.

Because p_i^* increases with d_1 and d_2 for $i \in \{1,2\}$ (from Proposition 1) and $p_i^* \to \infty$ as $d_1 \to \frac{1}{2\,\lambda}$ from (10), there exists a unique $d_1 < \frac{1}{2\,\lambda}$ that ensures equilibrium prices are those identical prices that leave the marginal consumer with zero surplus. From (10) and (43), this value of d_1 is determined by:

$$p_i^* = \frac{[\,1-\lambda\, d_1\,]\,[\,L-\Delta\,]\,[\,L+\Delta\,]}{L\,[\,1-2\,\lambda\, d_1\,]-\Delta} = v_1 - \frac{1}{2}\,[\,L-\Delta\,]$$

$$\Leftrightarrow\ d_1 = d_1^w = \left[\frac{1-\frac{\Delta}{L}}{2\,\lambda}\right]\frac{2\,[\,v_1-L\,]-L-\Delta}{2\,[\,v_1-L\,]+\Delta+\frac{\Delta^2}{L}} < \frac{1}{2\lambda}\,.$$

Finally, from (8), if both firms innovate, then welfare (not counting innovation costs) is

$$\widetilde{W}_{12} = v_1\, q_1^* - \frac{1}{2}\, q_1^{*2} + v_2\, q_2^* - \frac{1}{2}\, q_2^{*2} = v_1\, q_1^* - \frac{1}{2}\, q_1^{*2} + v_2\,[\,L-q_1^*\,] - \frac{1}{2}\,[\,L-q_1^*\,]^2$$

$$= -\,\Delta\, q_1^* - \frac{1}{2}\, q_1^{*2} + v_2\, L - \frac{1}{2}\,[\,L-q_1^*\,]^2\,. \quad (44)$$

The maximum feasible value of $\widetilde{W}_{12}$ is

$$\widetilde{W}_{12} = v_2\, L - \frac{1}{4}\left[\,2\,L\,\Delta - \Delta^2 + L^2\,\right],\ \text{ achieved via }\ q_1^* = \frac{1}{2}\,[\,L-\Delta\,]$$

because, from (44):

$$\frac{\partial \widetilde{W}_{12}}{\partial q_1^*} = -\,\Delta - q_1^* + L - q_1^* = 0 \quad\Rightarrow\quad q_1^* = \frac{1}{2}\,[\,L-\Delta\,]$$

and

$$\widetilde{W}_{12}\Big|_{q_1^*=\frac{L-\Delta}{2}} = -\,\Delta\,\frac{1}{2}\,[\,L-\Delta\,] - \frac{1}{8}\,[\,L-\Delta\,]^2 + v_2\, L - \frac{1}{2}\left(\frac{1}{2}\,[\,L+\Delta\,]\right)^2$$

$$= \; v_2\, L - \frac{1}{4}\left[\, 2\, L\, \Delta - \Delta^2 + L^2 \,\right]. \quad \blacksquare$$

Proof of Lemma 3

(2) and (3) imply that industry profit is:

$$\Pi \;=\; \frac{p_1}{2}\,[\, L + v_1 - v_2 + p_2 - p_1 \,] + \frac{p_2}{2}\,[\, L + v_2 - v_1 + p_1 - p_2 \,]. \tag{45}$$

Since $v_1 > 2\,L$ and $v_2 > 2\,L$ from Assumption 1, Π is highest when both firms produce positive output and all consumers purchase a unit of the product. Therefore, from (21) and (45), the value of p_1 that maximizes Π is determined by:

$$\frac{d\Pi}{dp_1} \;=\; \frac{\partial \Pi}{\partial p_1} + \frac{\partial \Pi}{\partial p_2}\frac{\partial p_2}{\partial p_1} \;=\; \frac{\partial \Pi}{\partial p_1} - \frac{\partial \Pi}{\partial p_2} \;=\; 0$$

$$\Leftrightarrow \quad L + v_1 - v_2 + 2\,p_2 - 2\,p_1 - (L + v_2 - v_1 + 2\,p_1 - 2\,p_2) \;=\; 0$$

$$\Leftrightarrow \quad \overline{p}_1 \;=\; \frac{3\,v_1 + v_2 - 2\,L}{4} \;=\; v_1 - \frac{1}{2}\left[L - \frac{\Delta}{2} \right]. \tag{46}$$

(21) and (46) imply:

$$\overline{p}_2 \;=\; v_1 + v_2 - \frac{3\,v_1 + v_2 - 2\,L}{4} - L \;=\; v_2 - \frac{1}{2}\left[L + \frac{\Delta}{2} \right]. \tag{47}$$

From (2), the corresponding equilibrium outputs are:

$$\overline{q}_1 \;=\; \frac{1}{2}\,[\, L + v_1 - v_2 + \overline{p}_2 - \overline{p}_1 \,] \quad \text{and} \quad \overline{q}_2 \;=\; \frac{1}{2}\,[\, L + v_2 - v_1 + \overline{p}_1 - \overline{p}_2 \,]. \tag{48}$$

These outputs are positive and sum to L. Substituting $\bar{p}_1$ and $\bar{p}_2$ into (45) and simplifying provides

$$\overline{\Pi} \;=\; L\left[v_1 - \frac{1}{2}\,(L - \Delta) \right] + \frac{\Delta^2}{8} \;=\; \Pi^w + \frac{\Delta^2}{8}.$$

Substituting $\overline{p}_1$ and $\overline{p}_2$ into (36) and simplifying provides the values of $\bar{d}_1$ and $\bar{d}_2$ specified in (14). $\blacksquare$

Proof of Observation 1.

From (10) and (18), if $d_2 \;=\; d_2^w\,(d_1)$, then:

$$k_2^w \;=\; \frac{v_1}{2}\,[\, L - \Delta \,] - \frac{\left[\, \frac{1}{2}\,(L - \Delta) \,\right]^2}{2} + \frac{v_2}{2}\,[\, L + \Delta \,] - \frac{\left[\, \frac{1}{2}\,(L + \Delta) \,\right]^2}{2} - v_1 + \frac{1}{2}$$

$$= \; \frac{L}{2}\,[\, v_1 + v_2 \,] + \frac{1}{4}\left[\, \Delta^2 - L^2 \,\right] - v_1 + \frac{1}{2} \;=\; k_2^{w*}. \tag{49}$$

If $m \;=\; 0$, then $\widehat{k}_2 \;=\; \min\left\{ \pi_2^*, \bar{k}_2 \right\}$ from (16). Furthermore, (16), (19), and integration by parts provide:

$$k_1^w(\widehat{k}_2) - \hat{k}_1 \;=\; G(\widehat{k}_2)\,\widetilde{W}_{12} + \Big[1 - G(\widehat{k}_2)\Big]\,\widetilde{W}_1 - \int_0^{\widehat{k}_2} k_2\,dG(k_2)$$

$$-\;\; G(\widehat{k}_2)\,\pi_1^* - \Big[1 - G(\widehat{k}_2)\Big]\,\pi_1^M$$

$$=\; G(\widehat{k}_2)\,\widetilde{W}_{12} - G(\widehat{k}_2)\Big[\hat{k}_2 + \pi_1^*\Big] + \int_0^{\widehat{k}_2} G(k_2)\,dk_2 + \Big[1 - G(\widehat{k}_2)\Big]\Big[\widetilde{W}_1 - \pi_1^M\Big]$$

$$\geq\; G(\widehat{k}_2)\Big[\widetilde{W}_{12} - (\pi_2^* + \pi_1^*)\Big] + \int_0^{\widehat{k}_2} G(k_2)\,dk_2 \;>\; 0. \quad \blacksquare$$

Proof of Observation 2.

If $v_1 = v_2 \equiv v$ and $d_1 = d_2 = m = 0$, then $q_1^* = q_2^* = L/2$ from (10), and $\widehat{k}_2 = \pi_2^* = p_2^*\, q_2^*$ from (16). Therefore, from (18):

$$k_2^w - \hat{k}_2 \;=\; \widetilde{W}_{12} - \widetilde{W}_1 - \hat{k}_2 \;=\; v\,q_1^* - \frac{q_1^{*2}}{2} + v\,q_2^* - \frac{q_2^{*2}}{2} - v + \frac{1}{2} - p_2^*\,q_2^*$$

$$=\; v\,[\,L - 1\,] - \frac{L^2}{4} + \frac{1}{2} - p_2^*\,\frac{L}{2}\,. \tag{50}$$

Furthermore, from (6):

$$p_2^* \;=\; \frac{1}{3}\,[\,3\,L + \Delta\,] \quad\Rightarrow\quad p_2^* \;=\; L \quad \text{when } v_1 \;=\; v_2\,. \tag{51}$$

(50) and (51) imply:

$$k_2^w - \hat{k}_2 \;=\; v\,[\,L - 1\,] - \frac{L^2}{4} + \frac{1}{2} - \frac{L^2}{2} \;=\; v\,[\,L - 1\,] + \frac{1}{4}\,\big[\,2 - 3\,L^2\,\big]\,. \tag{52}$$

(52) implies: (i) $k_2^w - \hat{k}_2 \;=\; -\,1/4 \;<\; 0$ if $L = 1$; and (ii) $k_2^w - \hat{k}_2 \;>\; 0$ if $L > 1$ and $v > \frac{1}{4}\left[\frac{3\,L^2 - 2}{L - 1}\right]$. $\blacksquare$

Proof of Proposition 2.

If (20) holds, then from Lemma 2, there exist a d_1 and $d_2 = d_2^w(d_1)$ such that

$$\frac{k_1^o - \pi_1^M}{G(k_2^o)} + \pi_1^M + k_2^o \;=\; \pi_1^* + \pi_2^*\,. \tag{53}$$

$$\text{Let} \quad \lambda\,m \;=\; \pi_2^* - k_2^o \quad\Rightarrow\quad \pi_2^* - \lambda\,m \;=\; k_2^o\,. \tag{54}$$

(53) and (54) imply:

$$\pi_1^* + \lambda\,m \;=\; \frac{k_1^o - \pi_1^M}{G(k_2^o)} + \pi_1^M \quad\Rightarrow\quad G(k_2^o)\,[\,\pi_1^* + \pi_2^* - k_2^o\,] + [\,1 - G\,(k_2^o)\,]\,\pi_1^M \;=\; k_1^o\,. \tag{55}$$

(16) and (53) – (55) imply $\hat{k}_2 = k_2^o$ and

$$\hat{k}_1 = G(\hat{k}_2)\,[\,\pi_1^* + \lambda\, m\,] + \Big[\,1 - G(\hat{k}_2)\,\Big]\,\pi_1^M$$
$$= G(k_2^o)\,[\,\pi_1^* + \pi_2^* - k_2^o\,] + [\,1 - G\,(k_2^o)\,]\,\pi_1^M \;=\; k_1^o\,. \tag{56}$$

Thus, the identified values of d_1, $d_2^w(d_1)$, and m induce the first-best outcome.

Finally, since $\lambda\, d_1 < 1$, it is apparent from (9) that $d_2 \gtreqless d_1$ as $\Delta \gtreqless 0$. ■

Proof of Corollary 1.

$G(k_2^o) = 1$ when $k_2^o = \bar{k}_2$. Therefore, since $k_1^o \le \bar{k}_1$, (20) holds if $\bar{k}_1 + \bar{k}_2 \le \Pi^w$. From Lemma 2, this inequality holds if:

$$\bar{k}_1 + \bar{k}_2 \;\le\; v_1\, L - \frac{L^2}{2} + \frac{L}{2}\,[\,v_2 - v_1\,] \;=\; \frac{L}{2}\,[\,v_1 + v_2 - L\,]\,.$$

The conclusions regarding $\bar{k}_1$, $\bar{k}_2$, v_1, and v_2 are apparent. The conclusion regarding L holds because $\frac{\partial\left(\frac{L}{2}\,[\,v_1 + v_2 - L\,]\right)}{\partial L} = v_1 + v_2 - 2\,L > 0$, from Assumption 1. ■

Proof of Corollary 2.

Suppose, to the contrary, there is a policy with $d_2 = d_2^w\,(d_1)$, $\hat{k}_1 = k_1^o$, and $\hat{k}_2 = k_2^o$. Then, since (20) does not hold:

$$k_1^o + k_2^o\, G\,(k_2^o) - [\,1 - G\,(k_2^o)\,]\,\pi_1^M \;>\; [\,\pi_1^* + \pi_2^*\,]\, G\,(k_2^o)$$
$$\Leftrightarrow \quad k_1^o \;>\; [\,\pi_1^* + \lambda\, m^*\,]\, G\,(k_2^o) + [\,1 - G\,(k_2^o)\,]\,\pi_1^M \;=\; \hat{k}_1\,,$$

contradicting $\hat{k}_1 = k_1^o$. ■

Proof of Proposition 3.

(i) From (21), prices that leave the marginal consumer with zero surplus (and so maximize industry profit) are determined by

$$p_2 \;=\; v_1 + v_2 - p_1 - L \quad \Leftrightarrow \quad \frac{1}{2}\,[\,L - \Delta + p_2 - p_1\,] \;=\; v_1 - p_1\,. \tag{57}$$

(2) and (57) imply:

$$q_1 \;=\; v_1 - p_1\,; \quad q_2 \;=\; L + p_1 - v_1\,; \;\text{ and}$$
$$\Pi \;=\; p_1\,[\,v_1 - p_1\,] + [\,v_1 + v_2 - p_1 - L\,]\,[\,L + p_1 - v_1\,]\,. \tag{58}$$

Because $\hat{k}_2 = \pi_2 - \lambda\, m$ from (16):

$$\pi_1 + \lambda\, m \;=\; \pi_1 + \pi_2 - \hat{k}_2 \;=\; \Pi - \hat{k}_2\,. \tag{59}$$

(58) and (59) imply that the optimal policy can be identified by choosing p_1 and $\hat{k}_2$ to maximize W. (8) implies $\frac{\partial \widetilde{W}_{12}}{\partial \hat{k}_2} = 0$. Therefore, from (17), the optimal p_1 and $\hat{k}_2$ satisfy:

$$\frac{\partial W}{\partial p_1} = \left[G(\hat{k}_2)\,\widetilde{W}_{12} + \left[1 - G(\hat{k}_2)\right] \widetilde{W}_1 - \int_0^{\hat{k}_2} k_2\, dG(k_2) - \hat{k}_1 \right] \frac{\partial F(\hat{k}_1)}{\partial p_1}$$

$$+ \; F(\hat{k}_1) \left[\widetilde{W}_{12} - \widetilde{W}_1 - \hat{k}_2 \right] \frac{\partial G(\hat{k}_2)}{\partial p_1} + F(\hat{k}_1)\, G(\hat{k}_2)\, \frac{\partial \widetilde{W}_{12}}{\partial p_1}$$

$$= \left[k_1^w(\hat{k}_2) - \hat{k}_1 \right] \frac{\partial F(\hat{k}_1)}{\partial p_1} + F(\hat{k}_1) \left[k_2^w - \hat{k}_2 \right] \frac{\partial G(\hat{k}_2)}{\partial p_1}$$

$$+ \; F(\hat{k}_1)\, G(\hat{k}_2)\, \frac{\partial \widetilde{W}_{12}}{\partial p_1} \; = \; 0\,, \text{ and} \tag{60}$$

$$\frac{\partial W}{\partial \hat{k}_2} = F(\hat{k}_1) \left[\widetilde{W}_{12} - \widetilde{W}_1 - \hat{k}_2 \right] \frac{\partial G(\hat{k}_2)}{\partial \hat{k}_2}$$

$$+ \left[G(\hat{k}_2)\,\widetilde{W}_{12} + \left(1 - G(\hat{k}_2)\right) \widetilde{W}_1 - \int_0^{\hat{k}_2} k_2\, dG(k_2) - \hat{k}_1 \right] \frac{\partial F(\hat{k}_1)}{\partial \hat{k}_2}$$

$$= \left[k_1^w(\hat{k}_2) - \hat{k}_1 \right] \frac{\partial F(\hat{k}_1)}{\partial \hat{k}_2} + F(\hat{k}_1) \left[k_2^w - \hat{k}_2 \right] \frac{\partial G(\hat{k}_2)}{\partial \hat{k}_2} \; = \; 0\,. \tag{61}$$

The last equalities in (60) and (61) reflect (18) and (19).

(60) and (61) imply that if $\widetilde{p}_1$ and $\widetilde{k}_2$ solve (22) and (23) with $\widetilde{p}_2 = v_1 + v_2 - \widetilde{p}_1 - L$, and if $\widetilde{\Pi} < \overline{\Pi}$, then the optimal policy is $(\widetilde{d}_1, \widetilde{d}_2, \widetilde{m})$.

(ii) If $\widetilde{\Pi} > \overline{\Pi}$ at the values of p_1 and $\hat{k}_2$ that solve (22) and (23), then the candidate solution is not feasible because the identified level of industry profit exceeds the maximum feasible level. Formally, an omitted maximum industry profit constraint ($\widetilde{\Pi} \leq \overline{\Pi}$) is violated. Therefore, if $\widetilde{\Pi} \geq \overline{\Pi}$, the optimal prices are those that maximize industry profit, as specified in (13). These prices are induced via the values of $\overline{d}_i$ identified in (14). Thus, the optimal policy is $\left(\overline{d}_1, \overline{d}_2, \overline{m}\right)$, where $\hat{k}_2$ solves (23). ■

Proof of Lemma 4.

First, Assumption 1 ensures that $\widetilde{W}_1^z(k_1, k_2)$ and $\pi_1^{Mz}(k_1, k_2)$ are maximized when $p_1^M(k_1, k_2) = p_1^M = v - 1$. Furthermore, W^z is increasing in $\widetilde{W}_1^z(\cdot)$, and $B_1(k_1 \,|\, k_1)$ and $B_2(k_2 \,|\, k_2; k_1)$ are both non-decreasing in $\pi_1^{Mz}(k_1, k_2)$.

Therefore, if $p_1^M(k_1, k_2)$ differed from p_1^M for some $k_1 \in [0, \overline{k}_1]$ and $k_2 \in [0, \overline{k}_2]$ for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) = 0$, W^z could be increased without violating (28) by setting $p_1^M(k_1, k_2) = p_1^M$ for all $k_1 \in [0, \overline{k}_1]$ and $k_2 \in [0, \overline{k}_2]$ for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) = 0$.

Next, suppose that $q_i^z(k_1, k_2) \neq q_i(p_1(k_1, k_2), p_2(k_1, k_2))$ as specified in equation (2) for some $i \in \{1,2\}$ and for some $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$ for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$ at a solution to [P-Z]$'$. Call this solution "Solution A1."

Now consider a candidate solution to [P-Z]$'$ ("Solution B1") that is identical to Solution A1 except that for $i = 1, 2$ and for all $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$ for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$: (i) $q_i^z(k_1, k_2) = q_i(p_1(k_1, k_2), p_2(k_1, k_2))$; and (ii) $T(k_1, k_2)$ is set to ensure $B_1(k_1 \,|\, k_1)$ and $B_2(k_2 \,|\, k_2; k_1)$ are both weakly higher under Solution B1 than under Solution A1. Assumption 1 ensures that the identified increases in $B_1(k_1 \,|\, k_1)$ and $B_2(k_2 \,|\, k_2; k_1)$ are feasible.

Next, Assumption 1 also ensures that $\widetilde{W}_{12}^z(k_1, k_2)$ is weakly higher under Solution B1 than under Solution A1 for all $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$, and strictly higher for some $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$.

Therefore, because W^Z is increasing in $\widetilde{W}_{12}^z(\cdot)$, W^Z is strictly higher under Solution B1 than under Solution A1. Consequently, Solution A1 cannot be a solution to [P-Z]$'$, so the proof is complete by contradiction. ■

<u>Proof of Lemma 5</u>.

Because $L \geq 1$, Lemma 4 implies that at a solution to [P-Z]$'$, for all k_1, k_2 for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$:

$$\widetilde{W}_{12}^z(k_1, k_2) - \widetilde{W}_1^z(k_1, k_2) - k_2 \;\geq\; 0\,. \tag{62}$$

If this were not the case, then W^Z could be increased without violating (28) by setting $\alpha_2(k_1, k_2) = 0$.

(62) implies that if $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) \in (0, 1)$ for some $k_1 \in [\,0, \overline{k}_1\,]$ and $k_2 \in [\,0, \overline{k}_2\,]$, then W^z could be (weakly) increased without violating (28) by setting $\alpha_2(k_1, k_2) = 1$ for the identified (k_1, k_2).

Now suppose there exists a solution to [P-Z]$'$ ("Solution A2") under which condition (iii) of the lemma does not hold for some (k_1, k_2) for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$. Among all $(\,p_1(k_1, k_2), p_2(k_1, k_2), T(k_1, k_2)\,)$ triples for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$ under Solution A2, let $(\,p_1^w(k_1, k_2), p_2^w(k_1, k_2), T^w(k_1, k_2)\,)$ denote the triple that secures the highest level of welfare. Consider a candidate solution to [P-Z]$'$ ("Solution B2") that is identical to Solution A2 except that for all k_1, k_2 for which $\alpha_1(k_1) > 0$ and $\alpha_2(k_1, k_2) > 0$:

$$p_i(k_1, k_2) \;=\; p_i^w(k_1, k_2) \text{ for } i \;=\; 1,\, 2 \quad \text{and} \quad T(k_1, k_2) \;=\; T^w(k_1, k_2)\,.$$

Solution B2 is a feasible solution to [P-Z]$'$ because Solution A2 is a feasible solution to [P-Z]$'$. Furthermore, expected welfare is at least as high under Solution B2 as under Solution A2. Therefore, Solution B2 is a solution to [P-Z]$'$.

Observe that for any $k_1 \in [\,0, \overline{k}_1\,]$ for which $\alpha_1(k_1) > 0$ at a solution to [P-Z]$'$:

$$\int_0^{\overline{k}_2} \{\, \alpha_2(k_1, k_2)\, [\, \widetilde{W}_{12}^z(k_1, k_2) - k_2\,] + [\,1 - \alpha_2(k_1, k_2)\,]\; \widetilde{W}_1^z(k_1, k_2)\,\}\, dG(k_2) \;\geq\; k_1\,. \tag{63}$$

Otherwise, W^Z could be increased without violating (28) by setting $\alpha_1(k_1) = 0$. If (63) holds as a strict inequality and $\alpha_1(k_1) \in (0,1)$ for some k_1 at a feasible solution to [P-Z]$'$, then W^Z could be strictly increased without violating (28) by setting $\alpha_1(k_1) = 1$. If (63) holds as an equality and $\alpha_1(k_1) \in (0,1)$ for some k_1 at a feasible solution to [P-Z]$'$, then W^Z would not be reduced and (28) would not be violated if $\alpha_1(k_1)$ were set equal to 1. Therefore, $\alpha_1(k_1) = 0$ or $\alpha_1(k_1) = 1$ at a solution to [P-Z]$'$.

Suppose $\alpha_1(k_1') = 0$ and $\alpha_1(k_1'') > 0$ for some $0 \leq k_1' < k_1'' \leq \overline{k}_1$ at a solution to [P-Z]$'$. Then (63) implies:

$$\int_0^{\overline{k}_2} \{ \alpha_2(k_1'', k_2) [\widetilde{W}_{12}^z(k_1'', k_2) - k_2] + [1 - \alpha_2(k_1'', k_2)] \, \widetilde{W}_1^z(k_1'', k_2) \} \, dG(k_2) \ \geq \ k_1'' \ > \ k_1' \, . \tag{64}$$

(64) implies that W^Z would increase (and (28) would not be violated) if $Z(k_1', k_2)$ were replaced by $Z(k_1'', k_2)$ for all $k_2 \in [0, \overline{k}_2]$. Therefore, by contradiction, $\alpha_1(k_1)$ must be nondecreasing in k_1 at a solution to [P-Z]$'$.

Suppose that for some $k_1 \in [0, \tilde{k}_1]$, $\alpha_2(k_1, k_2') = 0$ and $\alpha_2(k_1, k_2'') > 0$ for some $0 \leq k_2' < k_2'' \leq \overline{k}_2$ at a solution to [P-Z]$'$. Then (62) implies:

$$\widetilde{W}_{12}^z(k_1, k_2'') - \widetilde{W}_1^z(k_1, k_2'') \ > \ k_2'' \ > \ k_2' \, . \tag{65}$$

(65) implies that W^Z would increase (and (28) would not be violated) if $Z(k_1, k_2')$ were replaced by $Z(k_1, k_2'')$. Therefore, by contradiction, $\alpha_2(k_1, k_2)$ must be nondecreasing in k_2 for all $k_1 \in [0, \tilde{k}_1]$ at a solution to [P-Z]$'$. ■

<u>Proof of Proposition 4.</u>

Lemmas 4 – 5 imply that if $\tilde{k}_2 \in (0, \overline{k}_2)$, then for all $k_1 \in [0, \tilde{k}_1]$ and $k_2 \in [0, \tilde{k}_2]$:

$$\pi_2^*(p_1(k_1, k_2), p_2(k_1, k_2)) - T(k_1, k_2) - \tilde{k}_2 \ = \ 0 \, . \tag{66}$$

(66) implies that for all $k_1 \in [0, \tilde{k}_1]$ and $k_2 \in [0, \tilde{k}_2)$:

$$\pi_2^*(p_1(k_1, k_2), p_2(k_1, k_2)) - T(k_1, k_2) - k_2 \ > \ 0 \, . \tag{67}$$

Similarly, if $\tilde{k}_1 \in (0, \overline{k}_1)$, then for all $k_1 \in [0, \tilde{k}_1]$ and $k_2 \in [0, \tilde{k}_2]$:

$$[\pi_1^*(p_1(k_1, k_2), p_2(k_1, k_2)) + T(k_1, k_2)] \, G(\tilde{k}_2) + \pi_1^M [1 - G(\tilde{k}_2)] \ = \ \tilde{k}_1 \, . \tag{68}$$

(68) implies that for all $k_1 \in [0, \tilde{k}_1)$ and $k_2 \in [0, \tilde{k}_2]$:

$$[\pi_1^*(p_1(k_1, k_2), p_2(k_1, k_2)) + T(k_1, k_2)] \, G(\tilde{k}_2) + \pi_1^M [1 - G(\tilde{k}_2)] - k_1 \ > \ 0 \, . \tag{69}$$

(67) and (69) imply that (26) and (27) are satisfied at a solution to [P-Z]$'$. Therefore, a solution to [P-Z]$'$ is a solution to [P-Z].

The remainder of the proof follows from the discussion in the text. ■

Appendix B

This Appendix presents additional numerical solutions that illustrate the nature and the relative performance of the optimal linear rule.

Table B1 illustrates how outcomes change under the optimal linear rule as the firms' expected innovation costs change in the baseline setting. The table reflects the conclusion in Proposition 2 and Corollary 1 that the optimal linear rule becomes more likely to ensure the first-best outcome as the maximum possible innovation costs decline, *ceteris paribus*. Table B1 also illustrates that the optimal linear rule can secure the first-best outcome whether the distributions of the firms' innovation costs are identical or distinct. In addition, the table indicates that the lump sum component of the optimal payment from firm 2 to firm 1 (m^*) tends to increase as firm 1's expected innovation cost increases relative to firm 2's expected innovation cost, *ceteris paribus*.

$\overline{k}_1$	$\overline{k}_2$	d_1^*	d_2^*	m^*	W^*	W^{UE}	W^{LP}	W^{FB}
3	3	0.63	0.63	0	9.69	8.63	7.56	9.69
3	5	0.75	0.75	-2	8.69	7.6	7.22	8.69
5	3	0.75	0.75	2	8.69	5.18	6.28	8.69
5	5	0.81	0.81	0	7.69	6.59	6.22	7.69
5	8	0.81	0.81	-1.3	6.52	5.81	5.58	6.52
8	5	0.84	0.84	1.88	5.81	3.74	3.89	6.19
8	8	0.84	0.84	0.69	4.60	3.69	3.68	5.02

Table B1. The Effects of Changing Innovation Costs

Table B2 illustrates how the optimal linear rule and welfare change as the breadth of the market (L) changes following firm 2's innovation in the baseline setting. As L increases, welfare increases due to the increased consumer demand for the firms' products. In addition, W^*/W^{UE} and W^*/W^{LP}, measures of the relative performance of the optimal linear rule, increase due to the increase in the maximum feasible level of welfare as the market expands.

L	d_1^*	d_2^*	m^*	W^*	W^{UE}	W^{LP}	W^{FB}
1.0	0	0	0.50	4.51	4.48	4.50	4.51
1.6	0.84	0.84	0.42	6.40	5.51	5.54	6.4
1.8	0.81	0.81	0	7.69	6.59	6.22	7.69
2.0	0.81	0.81	0	9.0	6.6	7.0	9.0

Table B2. The Effects of Changing Market Breadth.

Table B3 illustrates the effects of changing the probability that firm 2 ultimately will be found to have infringed firm 1's patent in the baseline setting. As this probability (λ) increases, the optimal values of d_1^* and d_2^* decline because the increased likelihood of a damage payment implies that desired incentives can be secured with a smaller payment. Observe that the optimal values of d_1^* and d_2^* in Table B3 exceed 1 when λ is sufficiently small. Therefore, a requirement that $d_1^* \leq 1$ or $d_2^* \leq 1$ (perhaps reflecting limited liability considerations) can preclude the optimal (restricted) linear rule from securing the first-best outcome when λ is small.

λ	d_1^*	d_2^*	m^*	W^*	W^{UE}	W^{LP}	W^{FB}
0.30	1.35	1.35	0	7.69	6.05	6.13	7.69
0.50	0.81	0.81	0	7.69	6.59	6.22	7.69
0.70	0.58	0.58	0	7.69	5.78	6.21	7.69

Table B3. The Effects of Changing the Likelihood of Patent Infringement.

References

Amici Curiae Brief. (2016). Supporting Petitioners before the Supreme Court of the United States in *Samsung Electronics Co., Ltd., et al., Petitioners v. Apple Inc.*, Case No. 15-777 (http://www.scotusblog.com/wp-content/uploads/2016/01/7589141_1_Tech-companies-cert-amicus-brief-15-777.pdf).

Anton, J. and D. Yao. (2007). "Finding "Lost" Profits: An Equilibrium Analysis of Patent Infringement Damages," *Journal of Law, Economics, and Organization*, 23: 186-207.

Bessen, J. and E. Maskin. (2009). "Sequential Innovation, Patents, and Imitation," *RAND Journal of Economics*, 40: 611-35.

Bessen, J. and M. Meurer. (2008). *Patent Failure: How Judges, Bureaucrats, and Lawyers Put Innovators at Risk*. Princeton, NJ: Princeton University Press.

Chang, H. (1995). "Patent Scope, Antitrust Policy and Cumulative Innovation," *RAND Journal of Economics*, 26: 34-57.

Choi, J. P. (1998). "Patent Litigation as an Information Transmission Mechanism," *American Economic Review*, 88: 1249-1263.

Choi, J. P. (2009). "Alternative Damage Rules and Probabilistic Intellectual Property Rights: Unjust Enrichment, Lost Profits and Reasonable Royalty Remedies," *Information Economics and Policy*, 21: 145-157.

Farrell, J. and C. Shapiro. (2008). "How Strong are Weak Patents?" *American Economic Review*, 98: 1347-1369.

Frank, R. and D. DeFranco. (2000-2001). "Patent Infringement Damages: A Brief Summary," *Federal Circuit Bar Journal*, 10: 281-297.

Gallini, N. (1992). "Patent Policy and Costly Imitation," *Rand Journal of Economics*, 23: 52-63.

Gilbert, R. and C. Shapiro. (1990). "Optimal Patent Length and Breadth," *RAND Journal of Economics*, 21: 106-112.

Green, J. and S. Scotchmer. (1995). "On the Division of Profit in Sequential Innovation," *RAND Journal of Economics*, 26: 20-33.

Henry, M. and J. Turner. (2010). "Patent Damages and Spatial Competition," *Journal of Industrial Economics*, 58: 279-305.

Kamien, M. and Y. Tauman. (1986). "Fees Versus Royalties and the Private Value of a Patent," *Quarterly Journal of Economics*, 101: 471–491.

Katz, M. and C. Shapiro. (1986). "How to License Intangible Property," *Quarterly Journal of Economics*, 101: 567-589.

Klemperer, P. (1990). "How Broad Should the Scope of Patent Protection Be?" *RAND Journal of Economics*, 21: 113-130.

Lemley, M. and C. Shapiro. (2005). "Probabilistic Patents," *Journal of Economic Perspectives*, 19: 75-98.

Murray, F., P. Aghion, M. Dewatripont, J. Kolev, and S. Stern. (2016). "Of Mice and Academics: Examining the Effect of Openness on Innovation," *American Economic Journal: Economic Policy*, 8: 212-252.

Myerson, R. (1979). "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 47: 61-73.

Nordhaus, W. (1969). "An Economic Theory of Technological Change," *American Economic Review*, 59: 18-28

Sampat, B. and H. Williams. (2015). "How Do Patents Affect Follow-On Innovation? Evidence from the Human Genome," NBER Working Paper No. 21666.

Schankerman, M. and S. Scotchmer. (2001). "Damages and Injunctions in Protecting Intellectual Property," *RAND Journal of Economics*, 32: 199–220.

Scotchmer, S. (1991). "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 5: 29-41.

Shapiro, C. (2016). "Patent Remedies," *American Economic Review*, 106: 198-202.

Sparapani, T. (2015). "Apple v. Samsung and a Fight Over the Patents for Designs," *Forbes/Tech*, August 6 (http://www.forbes.com/sites/timsparapani/2015/08/06/apple-v-samsung-and-a-fight-over-the-patents-for-designs/#53faabe96e66).

Vascellaro, J. (2012). "Apple Wins Big in Patent Case: Jury Finds Samsung Mobile Devices Infringed Six Apple Patents, Awards $1.05 Billion in Damages," *Wall Street Journal: Tech*, August 25 (http://www.wsj.com/articles/SB1000087239639044435840457760981065808 2898).

Zhang, M. and K. Hylton. (2015). "Optimal Remedies for Patent Infringement," Boston University School of Law, Law & Economics Working Paper No. 15-53 (http://www.bu.edu/law/faculty/scholarship/workingpapers/2015.html).