Department of Economics
Working Paper Series

'Pay-for-delay with Follow-on Products'

Jorge Lemus [1]
Emil Temnyalov [2]

[1] University of Illinois Urbana-Champaign, USA
[2] University of Technology Sydney, Australia

Pay-for-delay with Follow-on Products

Jorge Lemus[*] and Emil Temnyalov[†]

May 20, 2019

Abstract

We study pay-for-delay settlements between a patent holder and a challenger when the patent holder can introduce follow-on products. We show that ignoring follow-on products biases the inferred competitive harm of pay-for-delay settlements (the "Actavis inference"). The reason is that patent invalidation triggers an earlier introduction of follow-on products which changes pay-for-delay negotiation's payoffs relative to the case of no follow-on products. When follow-on products are ignored, we show that an inference based on a reverse payment *over-estimates* patent strength. If parties cannot use payments (as in pure-delay settlements), follow-on products may push the parties to settle on an earlier entry date relative to the entry date negotiated in the absence of follow-on products, and litigation may arise in equilibrium.

JEL codes: D2, K2, K4, L4, L13, O3.
Keywords: pay-for-delay, product hopping, evergreening, antitrust, litigation

1   Introduction

The patent system is founded on the premise that granting an exclusion right to an inventor for a limited period of time encourages innovation. The right to exclude rivals creates market power, so patents trade-off consumer-welfare losses and innovation incentives. In practice, whether a patent holder can implement an exclusion right is uncertain (Lemley and Shapiro, 2005). In the pharmaceutical industry, a generic firm can enter a market protected by a patent after demonstrating: (a) bio-equivalence of its product and a patented product; and (b) that either the generic product does not infringe on the original patent or that the original patent is invalid.[1] In response to an entry attempt by a generic firm, the patent holder (the brand firm) typically files a lawsuit claiming patent infringement. These lawsuits have often been resolved through "pay-for-delay" settlements, where the brand firm and the generic agree on an entry date for the generic (before the expiration of the patent) and the brand firm makes a payment to the generic (called a "reverse payment").[2]

Pay-for-delay settlements have been at the center of the intellectual property and antitrust debate. The anti-competitive aspect of these settlements is that they forestall competition by preserving patents that are unlikely to withstand reexamination (weak patents). Given that competition between two substitute products lowers total industry profit relative to monopoly profits, a brand firm and a generic entrant have incentives to preserve monopoly profits and share the rents.[3] This is particularly worrisome when the brand firm holds a weak patent, because pay-for-delay settlement in this case can generate large consumer-surplus losses. Ghili and Schmitt (2017) compare the consumer surplus generated by the pay-for-delay settlement for the drug Lamictal with the consumer surplus generated in two counterfactual scenarios: a settlement without direct payment (i.e., a pure-delay settlement) and litigation to final judgment. They find that a pure-delay settlement would have increased consumer surplus by $100 million, whereas litigation to final judgment would have increased consumer surplus by $1.3 billion.

The FTC has challenged settlements involving reverse payments under the suspicion of anti-competitive conduct (Hemphill, 2006). In *FTC vs Actavis* (2013), the Supreme Court con-

---

[1] This is part of the Hatch-Waxman Act of 1984. For more details, see, e.g., Bulow (2004).

[2] The first generic to file a petition to enter the market obtains six months of generic exclusivity, which facilitates pay-for-delay settlements (see Hemphill (2006) and Hemphill and Lemley (2011)).

[3] Jacobo-Rubio et al. (2017) estimate that brand firms value deterrence at $4.6 billion on average while generic entrants value the right to enter, on average, at $236.8 million.

cluded that reverse payments are illegal if they are used to avoid competition, but they are not per-se illegal. The Supreme Court called on U.S. courts to apply the rule of reason when dealing with pay-for-delay cases. Edlin et al. (2015) interpreted this ruling by proposing the *Actavis Inference*: a settlement that involves a large and otherwise unexplained payment, combined with delayed entry, supports a reasonable inference of harm to consumers from lessened competition.

Pay-for-delay settlements have been studied in a setting with a single product. Our contribution is to study pay-for-delay settlements in a setting with more than one product. Specifically, in our setting the brand firm can introduce a new product that can be a radical innovation or just a minor modification of an existing product. This is an important feature of the pharmaceutical industry, where brand firms routinely introduce follow-on products, a strategy that has been called "product hopping" or "evergreening" when the follow-on product is a minor modifications (e.g., a reformulation from a capsule to a tablet). Product hopping has recently attracted the attention of antitrust authorities. Cases involving product hopping include AstraZeneca reformulating Prilosec into Nexium, two drugs used to treat severe stomach acid-related conditions;[4] Abbott reformulating TriCor (a drug used to treat high triglyceride levels) from capsules to tablets;[5] Reckitt switching its branded formulation Suboxone (a drug to treat opiod addiction) from a sublingual tablet to a sublingual film;[6] Warner Chilcott switching Doryx (an acne medication) from tablets to capsules;[7] Actavis and Forest Laboratories switching Nameda (an Alzheimer's drug) from an immediate release formulation to an extended release formulation.[8]

The main economic force in our analysis is the strategic use of follow-on products to moderate the losses from patent invalidation. If the patent for the first product is invalidated, the patent holder introduces the second product earlier. This strategic effect changes the payoffs in a pay-for-delay negotiation relative to the case of no follow-on products. Consider a simplified version of our model to illustrate this economic trade-off. First, suppose there is only one product. Under patent protection the brand firm gets a profit π for one period. In this case, the patent holder risks π by going to litigation, i.e., the patent holder losses π if the patent is invalidated. Next, consider a setting where there is a second product protected

---

[4]For more details, see, e.g., Feldman and Frondorf (2016).

[5]Abbott Laboratories v. Teva Pharmaceuticals, 432 F. Supp. 2d 408 (D. Del. 2006).

[6]In Re: Suboxone Antitrust Litigation (201., 64 F. Supp. 3d 665, 681-83 (E.D. Pa. 2014)

[7]Mylan Pharmaceuticals v. Warner Chilcott, No. 12-3824 (E.D. Pa. June 13, 2013).

[8]New York ex rel. Schneiderman v. Actavis PLC (Namenda), 787 F.3d 638 (2d Cir. 2015)

by an ironclad patent that generates a profit of π^H for the patent holder. We assume the second product completely cannibalizes the sales of the first product, and that the patent holder prefers to introduce the two products sequentially rather than simultaneously, i.e., $\pi^H < \pi + \beta\pi^H$, where $\beta \in (0, 1)$ is the discount factor. In this case, if the first product's patent is invalidated, the patent holder loses π from the first product but introduces the second product earlier, which creates a gain of $(1 - \beta)\pi^H$. The key insight is that the invalidation of the first patent removes the cannibalization between the first and the second product, which triggers an earlier introduction of the follow-on product and changes the patent holder's negotiation incentives.

We study settlements that delay entry both when reverse payments are allowed and when they are not. First, when reverse payments are allowed, we show settlements involve a *lower* transfer from the brand firm to the generic relative to the transfer in a model without follow-on products. Thus, an antitrust inference based on this payment, such as the Actavis inference, *over-estimates* the strength of the patent if follow-on products are not accounted for properly. Second, in pure-delay settlements, we show that follow-on products could push the parties to settle on an earlier date of entry relative to the equilibrium settlement date in a setting without follow-on products. Thus, an inference based on the length of the delay of entry that ignores follow-on products may *under-estimate* the strength of the patent. Importantly, if the introduction of a follow-on product increases the brand firm's outside option significantly more than its settlement profits, the result reverses and an inference based on the length of the delay leads to an *over-estimation* of the patent strength. As a consequence, welfare-loss estimates based on a settlement framework that ignores the endogenous introduction of a new product may lead to biased results. Finally, we show that under pure-delay settlement, litigation may arise in equilibrium. In this case a weak patent for the first product or longer development time for a follow-on product both make settlement agreement more likely. The model provides a novel explanation for equilibrium litigation even when firms hold symmetric beliefs about their litigation prospects.

Related Literature. To the best of our knowledge, we are the first to provide a formal analysis of pay-for-delay settlements with follow-on products. Carrier (2010) was among the first to discuss pay-for-delay and product hopping in conjunction. Carrier (2011) presents the case study of Provigil, showing the anticompetitive harm that can result from the combination of pay-for-delay and product hopping. Gallasch (2016) argues that product hopping is facilitated by pay-for-delay settlements. Feldman and Frondorf (2016) describe three "gen-

erations" of pay-for-delay strategies. In Generation 1.0, brand firms pay cash directly to generics to delay their entry. In Generation 2.0, the brand firm uses side deals, rather than cash, to disguise the payment to generics. In Generation 3.0, brand firms exploit regulation in combination with pay-for-delay to prevent generics from entering the market. Similarly, Kesselheim and Darrow (2015) discuss current and emerging challenges with the Hatch-Waxman Act, whereas Hovenkamp and Lemus (2018) discuss how PTAB, a new venue to challenge the validity of a patent, has influenced firms' settlement decisions. Dickey et al. (2009) shows that some of the conclusions in the analysis of pay-for-delay settlements change after including real-world complexities such as litigation costs, risk aversion, asymmetric information, time discounting, and cash constraints. Our model incorporates several of these features (e.g., litigation costs and time discounting) in addition to follow-on product.

2 Pay-for-Delay and Sequential Products

There are two firms, B and G, competing in a continuous time model. From time 0 firm B, the brand firm, sells product $\mathcal{O}$ that is protected by a patent of strength $\theta \in [0, 1]$ and statutory length $T > 0$. The patent strength corresponds to the probability that the patent will be found valid and/or infringed in court. We assume that θ is common knowledge for both firms. Firm G manufactures a generic version of product $\mathcal{O}$. At time 0, firm G attempts to enter the market by claiming that the patent protecting $\mathcal{O}$ is either invalid or non-infringed. This leads to litigation. If firm G wins the lawsuit, then firms G and B earn flow duopoly profits π^d for a period Δ. After this period, free-entry drives profits to zero. If firm G loses the lawsuit, product $\mathcal{O}$ is sold exclusively by firm B until T, because no other generic firm will challenge the validity of the patent, and firm B gets monopoly profits π^m.[9] After product $\mathcal{O}$'s patent expires at time T, firms B and G make zero profit from selling product $\mathcal{O}$ and its generic version, respectively.

Besides selling product $\mathcal{O}$, firm B can introduce a follow-on product. By paying a cost K, firm B is able to introduce product $\mathcal{H}$ at any time after time τ, where $\Delta < \tau < T - \Delta$.[10] The

---

[9]This assumption is explained by the privilege enjoyed by the first generic challenger as part of the Hatch-Haxman legislation: It is the only generic firm to gain 180-days of exclusivity. Thus, no other generic firm has an incentive to challenge the patent, because in the best-case scenario the patent is invalidated, but then free-entry drives profits to zero.

[10]The assumption on τ is made for the sake of exposition.

parameter τ measures the complexity of developing product $\mathcal{H}$: the lower τ, the quicker firm B can introduce the new product. We assume that product $\mathcal{H}$ fully cannibalizes the sales of product $\mathcal{O}$ or any of its generic versions. This assumption reflects that either product $\mathcal{H}$ is superior to product $\mathcal{O}$, or that product $\mathcal{H}$ is a minor modification of $\mathcal{O}$ but the brand firm invests in marketing to effectively switch the market from $\mathcal{O}$ to $\mathcal{H}$ (product hopping). We assume that product $\mathcal{H}$ is protected by an ironclad patent, which implies that firm B obtains a flow profit of π^H and firm G zero profit after the introduction of product $\mathcal{H}$. At time 0, after the outcome of the lawsuit—i.e., after the court determines the validity of product $\mathcal{O}$'s patent, or after firms reach a settlement—firm B decides whether or not, and when, to introduce the new product. The decision of *when* to introduce this new version of the product is contingent on the litigation outcome. We assume that firms discount time at a common rate $r > 0$. For $T_1 \leq T_2$, we define

$$\delta(T_1, T_2) = \int_{T_1}^{T_2} e^{-rt} dt = \frac{e^{-rT_1} - e^{-rT_2}}{r},$$

which represents the discount rate between T_1 and T_2. A useful property of this discount rate is $\delta(T_1, T_2) = \delta(0, T_2) - \delta(0, T_1) = e^{-rs}\delta(T_1 - s, T_2 - s)$. We assume that firms are sequentially rational and there is perfect information, so the equilibrium concept we use is subgame perfect equilibrium.

Incentives to introduce the new product

We assume that introducing product $\mathcal{H}$ at time T is profitable. This allows us to focus on the effect of pay-for-delay on the timing of new product introduction rather on the effect of enabling new products. The introduction of product $\mathcal{H}$ is profitable under the following assumption.

Assumption 1 (Introducing product $\mathcal{H}$ at T is profitable)**.** $K < \pi^H \delta(T, 2T)$.

In principle, π^H may be larger or smaller than π^m. If $\pi^H > \pi^m$, firm B may have an incentive to introduce product $\mathcal{H}$ before T to replace the lower flow profit generated by product $\mathcal{O}$. For simplicity, we study the case where firm B experiences "replacement effect," so it has incentives to delay the introduction of $\mathcal{H}$ if the patent of product $\mathcal{O}$ is valid. For this to hold, we impose that when firm B's patent for product $\mathcal{O}$ prevails in court, then this firm introduces product $\mathcal{H}$ at time T rather than at time $t < T$, i.e.,

$$\delta(0, \tau)\pi^m + \delta(\tau, T + \tau)\pi^H \leq \delta(0, T)\pi^m + \delta(T, 2T)\pi^H \quad \Leftrightarrow \quad \pi^H(1 - e^{-rT}) \leq \pi^m \quad (1)$$

Assumption 2 (Deferred New Product Introduction). $\pi^H(1 - e^{-rT}) < \pi^m$.

Under Assumption 2, firm B delays the introduction of product $\mathcal{H}$ until the patent of the first product expires. Note that Assumption 2 holds when $\pi^H \leq \pi^m$ or when $\pi^H > \pi^m$ and r is sufficiently low (firms are sufficiently patient). If the patent of the first product is invalidated by the court, or if there is a settlement agreement that delays the entry of firm G, the *timing* of the introduction of the new product may change. In other words, the timing of product $\mathcal{H}$'s introduction depends on litigation/settlement outcome. If firm B prevails in litigation, then product $\mathcal{H}$ will be introduced at time T, right after the patent of product $\mathcal{O}$ expires (Assumption 2). If the patent of product $\mathcal{O}$ is invalidated, then firm B has incentives to anticipate the entry of $\mathcal{H}$, because receiving π^H early is better than nothing after time Δ, given that $\delta(T, 2T) < \delta(\tau, T + \tau)$. Thus, firm B has incentives to introduce product $\mathcal{H}$ as soon as possible (i.e., at time τ) after the patent for $\mathcal{O}$ has been invalidated.

Litigation Payoffs. An invalidation of product $\mathcal{O}$'s patent triggers firm B to introduce product $\mathcal{H}$ at time τ rather than at time T. Considering this difference in the timing of product $\mathcal{H}$'s introduction, the litigation payoff of firm G and firm B are, respectively,

$$L_G = (1 - \theta)\delta(0, \Delta)\pi^d - c_G, \tag{2}$$

$$L_B = \theta[\delta(0, T)\pi^m + \delta(T, 2T)\pi^H] + (1 - \theta)[\delta(0, \Delta)\pi^d + \delta(\tau, T + \tau)\pi^H] - c_B - K. \tag{3}$$

Similarly, we denote by $\bar{L}_B$ and $\bar{L}_G$ the payoff of firm B and firm G, respectively, in the absence of product $\mathcal{H}$:

$$\bar{L}_B = \theta\delta(0, T)\pi^m + (1 - \theta)\delta(0, \Delta)\pi^d - c_B. \tag{4}$$

$$\bar{L}_G = L_G. \tag{5}$$

Note that firm G's litigation payoff is not altered by products $\mathcal{H}$, while firm B's litigation payoff *increases* with follow-on products. We have the following result:

Lemma 1. *A follow-on product increases firm B's litigation payoff, i.e.*

$$L_B \geq \bar{L}_B.$$

Proof. Firm B's litigation payoff in a setting without new products, equals $\bar{L}_B$. Notice that

$$L_B = \bar{L}_B + \pi^H[\theta\delta(T, 2T) + (1 - \theta)\delta(\tau, T + \tau)] - K.$$

Given that $\delta(\tau, T + \tau) \geq \delta(T, 2T)$, and by Assumption 1, we have $L_B > \bar{L}_B$. $\qquad\square$

Firm G has a credible litigation threat if and only if $L_G \geq 0$, i.e., $c_G \leq (1-\theta)\delta(0,\Delta)\pi^d$. Otherwise, if $L_G < 0$, the game ends.[11] The generic incentive initiate litigation is higher when the patent protection product $\mathcal{O}$ is weaker patents (i.e., lower θ), and it is unaffected by the existence of product $\mathcal{H}$.

The firms' joint litigation payoff $L = L_B + L_G$ is given by

$$
\begin{aligned}
L &= \theta[\delta(0,T)\pi^m + \delta(T,2T)\pi^H] + (1-\theta)[\delta(0,\Delta)2\pi^d + \delta(\tau,T+\tau)\pi^H] - c_B - c_G - K \\
&= \underbrace{\theta\delta(0,T)\pi^m + (1-\theta)\delta(0,\Delta)2\pi^d - c_B - c_G}_{\text{Myopic P4D}} + \underbrace{\pi^H[\theta\delta(T,2T) + (1-\theta)\delta(\tau,T+\tau)] - K}_{\text{New Product}}
\end{aligned}
$$

We can decompose the joint litigation payoff into two terms: a myopic pay-for-delay term (Myopic P4D), which is the standard term when we do not consider follow-on products, and a novel term given by the introduction of a new product (New Product).

Patent invalidation through litigation distorts firms' joint surplus for two reasons: (1) it erodes monopoly profits during $[0,T]$; (2) it forces firm B to introduce product $\mathcal{H}$ earlier than it would otherwise do.

Settlement. Under Assumption 2, firms maximize their joint settlement surplus by delaying the entry of the new product until time T, to avoid the replacement effect. In this case, the firms' best agreement is to delay firm G's entry until T, which effectively prevents the entry of any generic version, since the introduction of $\mathcal{H}$ completely cannibalizes the sales of product $\mathcal{O}$ or any of its generics. Thus, settlement payoffs are

$$
\begin{aligned}
S_G &= x, & (6) \\
S_B &= \delta(0,T)\pi^m + \delta(T,2T)\pi^H - x - K, & (7)
\end{aligned}
$$

where $x \geq 0$ is a transfer paid by firm B to firm G. The joint settlement payoff is

$$
S = \delta(0,T)\pi^m + \delta(T,2T)\pi^H - K.
$$

The term $\delta(0,T)\pi^m$ corresponds to the value of preserving monopoly profits until T, which is the standard term in pay-for-delay. The term $\delta(T,2T)\pi^H - K$ (which is positive by Assumption 1) corresponds to a novel term coming from new product introduction.

Incentive to Settle. We define the incentive to settle (the bargaining surplus) as the difference between the joint payoff from settlement (according to a full delay) and the joint

---

[11]Throughout the paper, unless otherwise noted, we focus on credible litigation threats by the generic.

litigation payoff. We can write the incentive to settle as

$$I = \underbrace{(1-\theta)[\delta(0,T)\pi^m - \delta(0,\Delta)2\pi^d] + c_B + c_G}_{\text{Myopic P4D }(I_{P4D})} + \underbrace{\pi^H(1-\theta)[\delta(T,2T) - \delta(\tau, T+\tau)]}_{\text{New Product }(I_N)}. \quad (8)$$

It can be shown that under Assumption 2 we have $I \geq 0$, so firms always settle. In Equation 8, we decompose the incentive to settle into two terms. The first term, I_{P4D} is the incentive to settle in the standard pay-for delay framework, and the novel term I_N is due to the introduction of the follow-on innovation. This term is negative and can be written as

$$I_N = \pi^H(1-\theta)[\delta(T,2T) - \delta(\tau, T+\tau)] = -\pi^H(1-\theta)r\delta(0,T)\delta(\tau,T) < 0.$$

It is easy to see that this term is more negative for lower values of θ (weaker patents) and for lower values of τ (faster introduction of the new product).

2.1   Pay-for-Delay settlement transfers with follow-on products

A follow-on product *reduces* the bargaining surplus relative to the standard case of pay-for-delay, i.e., $I = I_{P4D} + I_N < I_{P4D}$. The reason is that when the patent of product $\mathcal{O}$ is invalidated firm B anticipates product $\mathcal{H}$'s introduction. This anticipation *increases* the litigation payoff of firm B, by reducing the *replacement effect* of introducing the new product earlier (Assumption 2), which puts B in a stronger bargaining position relative to the case of no follow-on product. The implied Nash-bargaining transfer from firm B to firm G (the reverse payment) is

$$x = (1-\theta)\delta(0,\Delta)\pi^d - c_G + \alpha(I_{P4D} + I_N), \quad (9)$$

where $\alpha \in [0,1]$ corresponds to firm G's bargaining power. We can write this transfer as $x = x_{P4D} + \alpha I_N < x_{P4D}$, where x_{P4D} is the implied transfer in a pay-for-delay in a framework that ignores follow-on products. Thus, if the anti-competitiveness of a pay-for-delay settlement is evaluated by the size of the reverse payment, ignoring follow-on products under-estimates the anticompetitive effect of the agreement. In other words, the implied patent strength inferred from a reverse payment is biased upwards.

Proposition 1. *If the size of a reverse payment is used to infer the quality of product $\mathcal{O}$'s patent, ignoring follow-on products leads to an* over-estimation *of this patent's quality.*

Proof. Without a follow-on product ($\pi^H = 0$) the inference of θ from Equation 9 is

$$x = (1-\theta)[\delta(0,\Delta)\pi^d + \alpha\delta(0,T)\pi^m - \alpha\delta(0,\Delta)2\pi^d] + \alpha c_B - (1-\alpha)c_G.$$

Then,

$$\theta_{P4D} = 1 - \frac{x + (1 - \alpha)c_G - \alpha c_B}{\delta(0, \Delta)\pi^d + \alpha\delta(0, T)\pi^m - \alpha\delta(0, \Delta)2\pi^d}.$$

When $\pi^H > 0$, with a follow-on product, the inference of θ from Equation 9 is

$$\theta_H = 1 - \frac{x + (1 - \alpha)c_G - \alpha c_B}{\delta(0, \Delta)\pi^d + \alpha\delta(0, T)\pi^m - \alpha\delta(0, \Delta)2\pi^d - \alpha\pi^H r\delta(0, T)\delta(\tau, T)} < \theta_{P4D}.$$

$\square$

Proposition 1 shows that ignoring the follow-on product may lead one to believe that the strength of the patent involved in the pay-for-delay settlement is *higher* than it actually is. In other words, the Actavis inference is biased when we ignore follow-on products. As an illustration of Proposition 1, we provide a numerical example with payoffs similar in magnitude to those in Ghili and Schmitt (2017).

Example 1. *Consider $\pi^m = \pi^H = \$2.23$ billion per year, $\pi^d = \$720$ million per year $c_B = c_G = \$50$ million, $\alpha = 0.5$, $\tau = 5$ years, $T = 20$ years, $\Delta = 6$ months, and $r = 0.03$. In this case, if we observe a transfer of $x = \$9$ billion and we ignore the follow-on products, we would then infer that $\theta_{P4D} = 0.47$. However, if we consider that firm B can introduce a follow-on product that generates a profit of $\pi^H = 2.23$ billion per year, after observing a transfer of $x = \$9$ billion we would infer that $\theta = 0.27$. The 20 percentage point difference in the estimation of patent strength corresponds to an overestimation of 74 percent.*

Figure 1 (left panel) shows the estimation of θ ignoring follow-on products (dashed red line) and incorporating them (solid blue line), for values of the settlement transfer x ranging between \$3 billion and \$12 billion.[12] Figure 1 (right panel) shows the percentage point difference in the inference of patent strength. The figure shows that ignoring follow-on products leads to an over-estimation of patent quality, and this bias is higher the higher the reverse payment.

2.2 Pure-delay settlements with follow-on products

What if reverse payments are forbidden? We study "pure-delay" deals, i.e., those in which firm B allows firm G to enter at some time $D \leq T$, as part of the settlement agreement. The

---

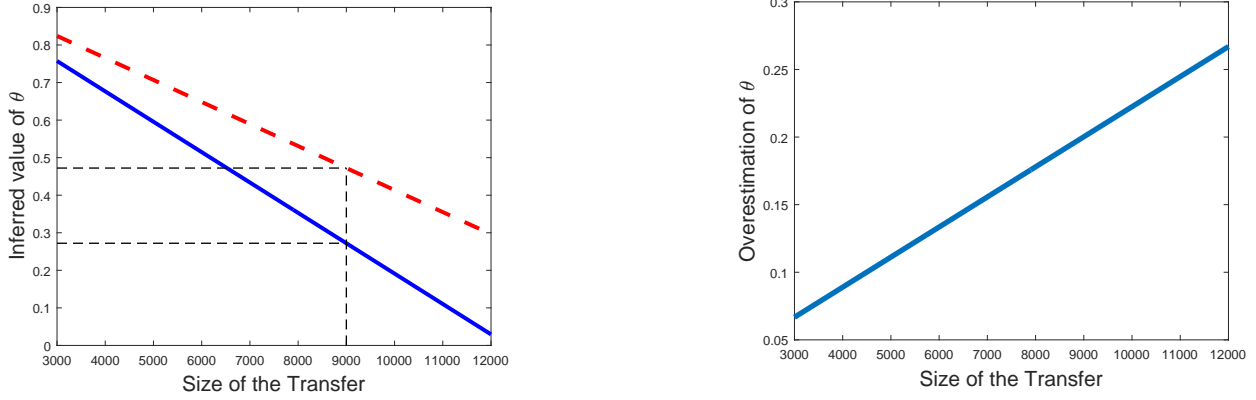[12]Ghili and Schmitt (2017) estimate a transfer between \$8 and \$12 billion.

Figure 1: Illustration of the bias in the estimation of patent strength by ignoring follow-on products. The dashed red line in the left panel corresponds to the inference when follow-on products are taken into account, whereas the solid blue line is the inference in a single-product setting.

economic tension is the following: if product $\mathcal{O}$'s patent is weak, a pure delay settlement must allow firm G to enter relatively soon. But in that case, firm B will be tempted to introduce the new product at τ rather than T, potentially reducing the rents of firm G. Therefore firm G would have to be compensated by entering even sooner.

If firm G enters at some time $D \leq \tau$, under the assumption that $\tau < T - \Delta$, firm B prefers to introduce the new product at τ rather than T if and only if

$$\pi^d \delta(\tau, T) + \delta(T, 2T)\pi^H < \pi^H \delta(\tau, \tau + T). \tag{10}$$

We assume that this condition always holds. Equivalently, this condition can be written as

Assumption 3 (Preemption). $\pi^d < \pi^H r \delta(0, T)$.

Under Assumption 2 and Assumption 3, firm G will not accept any delay longer than τ. Any delay longer than τ would result in zero profits for firm G. We now condider the equilibrium delay in pure-delay settlements with a follow-on product. Firm G's payoff from accepting a delay of $D \in [0, T]$ to enter is

$$S_G(D) = \begin{cases} \delta(D, \tau)\pi^d & \text{if } D \leq \tau, \\ 0 & D \geq \tau, \end{cases}$$

while firm B's payoff from that delay is

$$S_B(D) = \begin{cases} \delta(0, D)\pi^m + \pi^d \delta(D, \tau) + \pi^H \delta(\tau, \tau + T) & \text{if } D \leq \tau, \\ \delta(0, \tau)\pi^m + \pi^H \delta(\tau, \tau + T) & \text{if } D > \tau. \end{cases}$$

A delay D is feasible only if both firms unilaterally prefer that delay to litigation. Therefore, the maximum acceptable delay for firm G must be larger than the minimum acceptable delay for firm B. Any acceptable delay for firm G must be shorter than

$$D_G \equiv \max\{D : S_G(D) \geq L_G\}.$$

Given that $S_G(D)$ is decreasing and continuous, when $D \leq \tau$, D_G is the (unique) solution to

$$(1 - \theta)\delta(0, \Delta)\pi^d - c_G = \delta(D_G, \tau)\pi^d. \tag{11}$$

Similarly, the minimum acceptable delay for firm B, denoted by D_B, satisfies

$$D_B \equiv \min\{D : S_B(D) \geq L_B\}.$$

The firms' settlement and litigation payoffs in a single-product setting, i.e., a setting without product $\mathcal{H}$, are:

$$\bar{S}_B(D) = \delta(0, D)\pi^m + \delta(D, T)\pi^d,$$
$$\bar{S}_G(D) = \delta(D, T)\pi^d.$$

and $\bar{D}_B$ and $\bar{D}_G$ are defined, analogously to the case with follow-on products, by

$$\bar{D}_G \equiv \max\{D : \bar{S}_G(D) \geq \bar{L}_G\},$$
$$\bar{D}_B \equiv \min\{D : \bar{S}_B(D) \geq \bar{L}_B\}.$$

In Figure 2 we illustrate these thresholds for the parameter values in Example 1. Figure 2 (left panel) shows the maximum acceptable delay by the generic (solid blue) and the minimum acceptable delay by firm B (dashed red). The figure shows that $\bar{D}_G > \bar{D}_B$ for any θ such that the generic has a credible threat ($L_G \geq 0$). This implies that the firms always settle at some delay in the interval $[\bar{D}_B, \bar{D}_G]$. In contrast Figure 2 (right panel) shows that this is not longer true with follow-on products: for stronger patents (higher θ) the firms do not settlement because the minimum acceptable delay by the patent holder is larger than the maximum acceptable delay for the generic. Thus, with follow-on products, pure-delay settlements are more likely to occur when the underlying patent is weak (low θ).

The next proposition characterizes whether firms settle or pursue litigation in equilibrium.

Proposition 2. *When the patent holder can introduce a new product, a pure-delay settlement can be reached in equilibrium if and only if*

$$(1 - \theta)\delta(\Delta, \tau)\pi^m \geq \theta\delta(\tau, T)[\pi^m - \pi^H(1 - e^{-rT})] - c_B - c_G\left(\frac{\pi^m - \pi^d}{\pi^d}\right) - K.$$
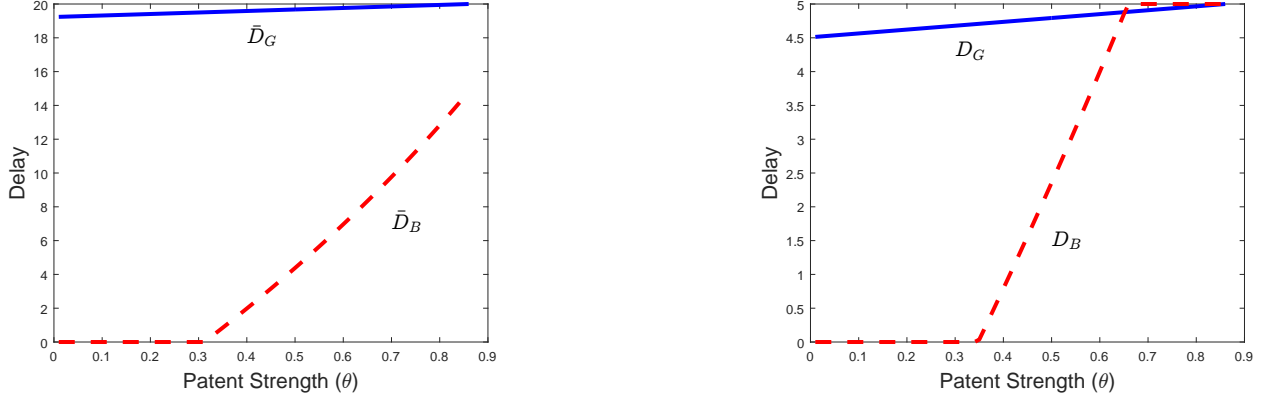
Otherwise the equilibrium features litigation on-path.

Figure 2: Maximum acceptable delay for the generic (solid blue lines) and minimum acceptable delay for the brand firm (dashed red lines). The left panel shows the case of no follow-on products and the right panel shows the case with follow on products.

Proof. To have settlement in equilibrium, we must have that B's payoff from a settlement at G's maximum acceptable delay must be larger than the outside option of litigation, i.e. $S_B(D_G) \geq L_B$. Using Equation 11 and the definition of S_B and L_B we get

$$\delta(0, D_G)\pi^m - c_G \geq \theta[\delta(0, T)\pi^m + \delta(T, 2T)\pi^H - \delta(\tau, T + \tau)\pi^H] - c_B - K$$

Using that $\delta(0, D_G) = \delta(0, \tau) - \delta(D_G, \tau) = \delta(0, \tau) - (1 - \theta)\delta(0, \Delta) + \frac{c_G}{\pi^d}$ and that $\delta(D_G, \tau) = \delta(0, \tau) - \delta(0, D_G)$, we get that settlement is incentive compatible if and only if the condition in the proposition holds. Finally, since a pure-delay settlement is always incentive compatible in the absence of follow-on products, it follows that litigation is less likely in that case. $\square$

Proposition 2 shows that litigation may be unavoidable as the result of follow-on products, whereas in the single-product setting a settlement is *always* incentive compatible. Thus, follow-on products provide a novel rationale for why costly litigation might occur in equilibrium.[13] Note that litigation arises in equilibrium for some values of θ when $c_B = c_G = 0$ and K is not too large, because the left-hand side of inequality Proposition 2 goes to zero as θ goes to one while the right-hand side of the inequality is strictly positive. Figure 3 illustrates the settlement/litigation boundary, defined by imposing equality in the condition in Proposition 2, for different product introduction times (τ) and patent strength (θ) and fixing other parameter values as in Example 1. First, for θ sufficiently low, a settlement is always incentive compatible. For intermediate values of patent strength, a smaller devel-

---

[13]The literature has discussed other forces such as asymmetric information (Bebchuk, 1984), divergence of beliefs (Priest and Klein, 1984), risk aversion (Heyes et al., 2004), or reputation (Che and Yi, 1993).

opment time of the new product induces litigation in equilibrium. Finally, notice that for $\theta \geq 0.86$ the generic does not have a credible litigation threat.
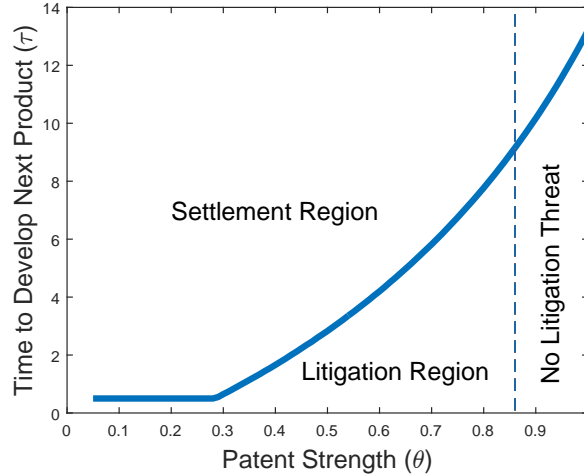

Figure 3: The settlement/litigation boundary for different product introduction times (τ) and patent strength (θ), for the same parameter values as in Example 1.

Corollary 1. *Lower patent strength facilitates settlements. An easy follow-on innovation (i.e., a smaller τ) may prevent settlement and may facilitate litigation in equilibrium.*

The corollary comes directly from comparative statics on the inequality in the condition in Proposition 2. Interestingly, the more "difficult" it is for firm B to introduce a new product, the more firm G is willing to tolerate a larger delay. Thus, a generic firm will not accept long delay if it believes that τ is small. For instance, if product $\mathcal{H}$ is a minor modification of the original product (product hopping), τ is presumably small. For short development time (low τ), a generic will accept a delay only when the underlying patent is weak (low θ). But if the original patent is strong and τ was small (as in the case of product hopping) litigation ensues. Why did the generics in the cases discussed by Carrier (2011) and Gallasch (2016) accepted a delay? One plausible explanation is that generic firms did not anticipate the patent holder's follow-on product.

Bargaining Solution for Pure Delay. Next, we focus on the case where settlement is feasible, i.e. the condition in Proposition 2 holds, and we examine the outcome of Nash bargaining over pure-delay settlements, which is the solution to the following problem:

$$\max_{D \in [D_B, D_G]} \quad [S_B(D) - L_B]^{(1-\alpha)} \cdot [S_G(D) - L_G]^{\alpha}. \tag{12}$$

In contrast, in the absence of follow-on products, the equilibrium pure-delay settlement is the solution to:

$$\max_{D \in [\bar{D}_B, \bar{D}_G]} \quad [\bar{S}_B(D) - \bar{L}_B]^{(1-\alpha)} \cdot [\bar{S}_G(D) - \bar{L}_G]^\alpha. \tag{13}$$

It is easy to see that follow-on products decreases firm G's settlement payoff for any given delay D, i.e. $S_G(D) \leq \bar{S}_G(D)$, and that firm G's outside option is unaffected, i.e., $L_G = \bar{L}_G$. On the other hand, it is clear that follow-on products increase firm B's settlement payoff for any given delay D, i.e. $S_B(D) \geq \bar{S}_B(D)$, and also they increase firm B's outside option, i.e. $L_B \geq \bar{L}_B$ (Lemma 1). The net effect of follow-on products on B's Nash settlement surplus with any delay D is ambiguous, depending on whether the increase in the settlement payoff or the outside option is relatively larger. The following proposition characterizes the Nash settlement outcome with and without follow-on products.

Proposition 3. *Follow-on products have ambiguous effect on the equilibrium delay of pure-delay settlement. If α is sufficiently small or if $[\theta \pi^H r \delta(0,T) - \pi^d]\delta(\tau,T) + K \geq 0$, then follow-on products induce shorter equilibrium delays relative to the setting with a single product.*

Proof. See Appendix. □

The intuition behind Proposition 3 is that a follow-on product worsens firm G's bargaining position but it can either improve or worsen firm B's barganinig position, because it improves both firm B's settlement and litigation payoffs. Delays are shorter when the improvement in the settlement payoff dominates, and longer when the improvement in the litigation payoff dominates, which may possibly even make settlement infeasible, if the improvement in B's litigation payoff is so large that settlement is no longer incentive compatible. An especially clear case is when firm B has all the bargaining power ($\alpha = 0$) and makes a take-it-or-leave-it offer to firm G. In this case, the best incentive-compatible offers that firm B can make with and without follow-on products are given by $S_G(D^*) = L_G$ and $\bar{S}_G(\bar{D}^*) = \bar{L}_G = L_G$, respectively. Since $\bar{S}_G(D) = S_G(D) + \delta(\tau,T)\pi^d \geq 0$, and $S_G(\cdot)$ is decreasing, it is clear that $\bar{D}^* \geq D^*$, i.e. a follow-on product leads to a shorter delay.

Corollary 2. *Follow-on products lead to shorter delays when firm B has all the bargaining power and makes a pure-delay take-it-or-leave-it offer to firm G.*

Figure 4 shows firm B and firm G's settlement and litigation payoffs with a follow-on product (solid blue line) and without follow-on products (dashed red line). The figure illustrates the

ambiguous effect of follow-on products on the equilibrium pure-delay settlements (when a settlement is incentive compatible). Figure 4 (left panel) shows a case where product $\mathcal{H}$ increases firm B's settlement payoff (S_B) relatively more than it increases firm B's litigation payoff (L_B). This implies that the minimum acceptable delay by firm B decreases, so the equilibrium delay is shorter. Figure 4 (right panel) shows the opposite case, where L_B increases relatively more than S_B, potentially leading to an increase in the equilibrium delay for some values of α. In each case, the relative bargaining power of the firms determine the equilibrium delay $\bar{D}^*$.

Finally, we study how follow-on products affect a regulator's inference about the strength of firm B's patent from an observed delay in a pure-delay settlement. Our analysis shows that ignoring follow-on products leads to a biased inference: in this case, conditional on any patent strength, equilibrium delay is often (but not always) shorter.

Corollary 3. *The value of patent strength inferred from a delay in a pure-delay settlement can be over or under estimated when follow-on products are ignored. When α is sufficiently small and $[\theta \pi^H r \delta(0, T) - \pi^d] \delta(\tau, T) + K \geq 0$, ignoring follow-on products under-estimates the value of θ.*
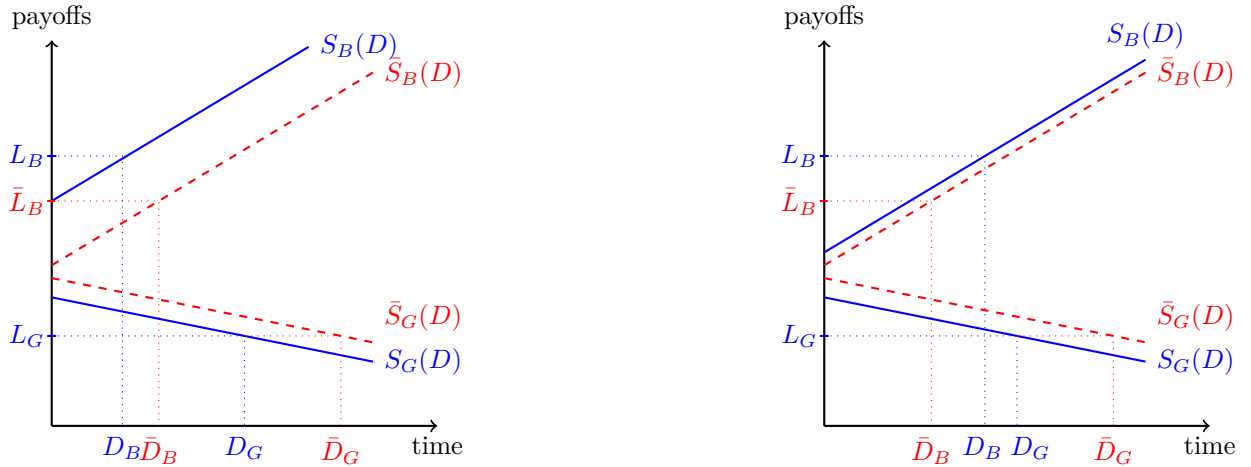

Figure 4: Pure-delay settlement and litigation payoffs for two sets of parameters. Payoffs with a follow-on product are in solid blue lines and payoffs without a follow-on product are in dashed red lines.

3 Policy Implications and Discussion

The first policy implication of our analysis is that ignoring follow-on products in pay-for-delay settlements biases antitrust inferences. First, when reverse payments are allowed, the strength of a patent inferred from the size of the reverse payment (the Actavis inference) will be *over-estimated.* Second, if reverse payments are forbidden and firms settle according to a pure-delay settlement, the strength of a patent inferred from the length of entry delay can be over- or under-estimated. When the brand firm has most of the negotiation power, however, equilibrium delays will be shorter with follow-on products.

When reverse payments are allowed, delay in generic entry is the same regardless of whether follow-on products exist or not. The only effect of follow-on products in this case is on the size of the reverse payment: the brand firm makes a lower payment to the generic firm. The delay in generic entry with follow-on products in the case of pure-delay settlements is ambiguous. On the one hand, if a pure-delay settlement agreement is feasible, follow-on products may push for an earlier introduction of the second product. This is good for consumers, even if the new product is a minor modification (product hopping) rather than a radical innovation, because that product would have been introduced at a later date anyway. But if pure-delay settlements are unfeasible and litigation ensues, then welfare may decrease because settling litigation provides some private and social benefits—savings in litigation costs, resolution of uncertainty, reduction of congestion in the court system. We show that weaker patents (lower θ) make settlement agreements more likely, and also that "quicker" development of follow-on products (lower τ) make settlement agreements less likely.

In our setting we assume that the brand firm always finds it profitable to introduce a follow-on product. This implies that a pay-for-delay settlement does not *enable* the patent holder to introduce a new product, but it rather changes the timing of introduction of this new product. In a different setting, pay-for-delay settlements may change a brand firm's incentive to introduce a new product, and enable follow-on products. As Carrier (2011) and Gallasch (2016) argue, this can be welfare reducing, if the follow-on product is a marginal improvement of the original drug (product hopping) that would have not be introduced otherwise.[14]

---

[14]Carrier (2011) and Gallasch (2016) argue that pay-for-delay and product hopping are a lethal combination because pay-for-delay enables the brand firm to "switch" to another product without having to fear any generic competition.

4 References

Bebchuk, Lucian Arye (1984) "Litigation and settlement under imperfect information," *The RAND Journal of Economics*, pp. 404–415.

Bulow, Jeremy (2004) "The gaming of pharmaceutical patents," *Innovation policy and the economy*, Vol. 4, pp. 145–187.

Carrier, Michael A (2010) "A Real-World Analysis of Pharmaceutical Settlements: The Missing Dimension of Product Hopping," *Fla. L. Rev.*, Vol. 62, p. 1009.

——— (2011) "Provigil: A Case Study of Anticompetitive Behavior," *Hastings Sci. & Tech. LJ*, Vol. 3, p. 441.

Che, Yeon-Koo and Jong Goo Yi (1993) "The role of precedents in repeated litigation," *Journal of Law, Economics, & Organization*, Vol. 9, pp. 399–424.

Dickey, Bret, Jonathan Orszag, and Laura Tyson (2009) "An Economic Assessment of Patent Settlements in the Pharmaceutical Industry," *Annals Health L.*, Vol. 19, p. 367.

Edlin, Aaron, Herbert Hovenkamp, Scott Hemphill, and Carl Shapiro (2015) "The Actavis Inference: Theory and Practice (Symposium: Drug Patent Settlements After Actavis)," *Rutgers University Law Review*, Vol. 67, p. 585.

Feldman, Robin and Evan Frondorf (2016) "Drug wars: a new generation of generic pharmaceutical delay," *Harv. J. on Legis.*, Vol. 53, p. 499.

Gallasch, Sven (2016) "A new dimension to EU pharma antitrust product hopping and unilateral pay for delay," *European Competition Journal*, Vol. 12, pp. 137–158.

Ghili, Soheil and Matt Schmitt (2017) "A Framework for Estimating Damages in Reverse Payment Cases," *Antitrust Law Journal*, Vol. 81, pp. 873–901.

Hemphill, C Scott (2006) "Paying for delay: Pharmaceutical patent settlement as a regulatory design problem," *NYUL Rev.*, Vol. 81, p. 1553.

Hemphill, C Scott and Mark A Lemley (2011) "Earning exclusivity: generic drug incentives and the Hatch-Waxman Act," *Antitrust Law Journal*, Vol. 77, p. 947.

Heyes, Anthony, Neil Rickman, and Dionisia Tzavara (2004) "Legal expenses insurance, risk aversion and litigation," *International Review of Law and Economics*, Vol. 24, pp. 107–119.

Hovenkamp, Erik and Jorge Lemus (2018) "Delayed entry settlements at the patent office," *International Review of Law and Economics*, Vol. 54, pp. 30–38.

Jacobo-Rubio, Ruben, John Turner, and Jonathan Williams (2017) "The Distribution of Surplus in the US Pharmaceutical Industry: Evidence from Paragraph (iv) Patent Litigation Decisions."

Kesselheim, Aaron S and Jonathan J Darrow (2015) "Hatch-Waxman turns 30: do we need a re-designed approach for the modern era," *Yale J. Health Pol'y L. & Ethics*, Vol. 15, p. 293.

Lemley, Mark A and Carl Shapiro (2005) "Probabilistic patents," *Journal of Economic Perspectives*, Vol. 19, pp. 75–98.

Priest, George L and Benjamin Klein (1984) "The selection of disputes for litigation," *The Journal of Legal Studies*, Vol. 13, pp. 1–55.

A   Appendix

Proof of Proposition 3

Proof. An interior solution to Equation 12 is characterized by

$$(1-\alpha)S'_B(D)[S_G(D) - L_G] + \alpha S'_G(D)[S_B(D) - L_B] = 0. \tag{14}$$

Also, notice that $S''_B(D) = -re^{-rD}(\pi^m - \pi^d)$ and $S''_G(D) = re^{-rD}\pi^d$, so $S''_B(D) + S''_G(D) < 0$. A sufficient condition for the objective function to be concave is

$$(1-\alpha)\frac{S''_B(D)}{S_B(D) - L_B} + \alpha\frac{S''_G(D)}{S_G(D) - L_G} \le 0,$$

which is satisfied for interior values of D as long as α is sufficiently small.

Analogously, an interior solution to Equation 13 is characterized by

$$(1-\alpha)\bar{S}'_B(D)[\bar{S}_G(D) - \bar{L}_G] + \alpha\bar{S}'_G(D)[\bar{S}_B(D) - \bar{L}_B] = 0. \tag{15}$$

Next, note that $S'_B(D) > 0$ and $\bar{S}'_B(D) > 0$, while $S'_G(D) < 0$ and $\bar{S}'_G(D) < 0$, and moreover $S_G(D) - L_G$ and $\bar{S}_G(D) - \bar{L}_G$ are decreasing in D, while $S_B(D) - L_B$ and $\bar{S}_B(D) - \bar{L}_B$ are increasing in D. Also, $S'_B(D) = \bar{S}'_B(D)$ and $S'_G(D) = \bar{S}'_G(D)$.

Furthermore, note that $S_G(D) - L_G \leq \bar{S}_G(D) - \bar{L}_G$. We also have that

$$[S_B(D) - L_B] - [\bar{S}_B(D) - \bar{L}_B] =$$
$$= \delta(\tau, T)(\pi^H - \pi^d) + \delta(T, \tau + T)\pi^H - \pi^H[\theta\delta(T, 2T) + (1 - \theta)\delta(\tau, \tau + T)] + K$$
$$= \pi^H[\delta(\tau, \tau + T) - \theta\delta(T, 2T) - (1 - \theta)\delta(\tau, \tau + T)] - \pi^d\delta(\tau, T) + K$$
$$= \pi^H\theta[\delta(\tau, \tau + T) - \delta(T, 2T)] - \pi^d\delta(\tau, T) + K$$
$$= [\theta\pi^H r\delta(0, T) - \pi^d]\delta(\tau, T) + K$$

By Assumption 3 this is always positive for θ sufficiently large. Moreover, since the π^H term is positive, a sufficient condition for the whole expression to be positive is that $K \geq \pi^d\delta(\tau, T)$. Thus if $[\theta\pi^H r\delta(0, T) - \pi^d]\delta(\tau, T) + K \geq 0$, we have

$$S_B(D) - L_B \geq \bar{S}_B(D) - \bar{L}_B, \tag{16}$$
$$S_G(D) - L_G \leq \bar{S}_G(D) - \bar{L}_G. \tag{17}$$

For the optimal $\bar{D}^*$ which solves Equation 15, we have

$$(1 - \alpha)\bar{S}'_B(\bar{D}^*)[\bar{S}_G(\bar{D}^*) - \bar{L}_G] + \alpha\bar{S}'_G(\bar{D}^*)[\bar{S}_B(\bar{D}^*) - \bar{L}_B] = 0,$$

and since $S'_B(\bar{D}^*) = \bar{S}'_B(\bar{D}^*)$, $S'_G(\bar{D}^*) = \bar{S}'_G(\bar{D}^*)$, $S_B(\bar{D}^*) - L_B \geq \bar{S}_B(\bar{D}^*) - \bar{L}_B$ and $S_G(\bar{D}^*) - L_G \leq \bar{S}_G(\bar{D}^*) - \bar{L}_G$, the optimality condition implies:

$$(1 - \alpha)S'_B(\bar{D}^*)[S_G(\bar{D}^*) - L_G] + \alpha S'_G(\bar{D}^*)[S_B(\bar{D}^*) - L_B] < 0.$$

Note that the $(1 - \alpha)$ term is positive, while the α term is negative. Next, since $S_G(D) - L_G$ is decreasing in D, and $S_B(D) - L_B$ is increasing in D, the optimal D^* which solves Equation 14 must decrease relative to $\bar{D}^*$. Hence, under the condition $[\theta\pi^H r\delta(0, T) - \pi^d]\delta(\tau, T) + K \geq 0$, for any α, the pure-delay settlement with follow-on innovation features a shorter delay: $D^* < \bar{D}^*$.

Finally, if α is sufficiently close to 0, the fact that $S_G(D) - L_G \leq \bar{S}_G(D) - \bar{L}_G$ implies that the solution to Equation 14 involves shorter delay than the solution to Equation 15, i.e. $\bar{D}^* < D^*$, with no additional conditions imposed the other parameters. $\qquad\square$