Appendix A: Proofs.

Proof of Proposition 1. Our continuous-time model allows us to analyze a recall decision using the derivative with respect to t_R.[A1] The condition of Case (A) can be obtained by using the minimum production quantity requirement (i.e., $s(t) = Nf_H(t) \leq N\gamma_0$). For the proof of Cases (B) and (C) below, we thus only consider the region of $\gamma + \lambda_H(I)F_H(t_I) > \gamma_0/[1 - F_H(t_I)]$.

Next, we start by finding the conditions for $\partial\pi_H/\partial t_R|_{t_R=t_I} > 0$ to identify when an delayed recall is always preferred. We first write out $\partial\pi_H/\partial t_R|_{t_R=t_I}$:

$$\frac{\partial\pi_H}{\partial t_I} = N[mf_H(T_R)\frac{dT_R}{\partial t_I} - (c_R + r_H)f_H(t_I)] = N[m\gamma_0\frac{dT_R}{dt_I} - (c_R + r_H)f_H(t_I)],$$

in which $dT_R/dt_I = [\gamma'(\overline{\tau}) - \lambda'(I)F_H(T_R)][1 - F_H(T_R)]/[f'_H(T_R)]$. As $f'_H(T_R) < 0$, then $\gamma'(\overline{\tau}) \geq \lambda'(I)F_H(T_R)$ implies a non-positive dT_R/dt_I, leading to a non-positive $\partial\pi_H/\partial t_I$, which indicates that an instant recall might be preferred by the firm,[A2] thus yielding a situation in Case (B): a firm may not delay a recall when $\gamma'(\overline{\tau}) \geq \lambda'(I)F_H(T_R)$.

However, if $dT_R/dt_I > 0$ (i.e., $\gamma'(\overline{\tau}) < \lambda'(I)F_H(T_R)$), the firm might consider implementing a delayed recall depending on its product margin-to-recall-cost ratio and/or the identification time of the product defect t_I. Using the definitions of $h(t_I) = f_H(t_I)/[\gamma_0(dT_R/dt_I)]$ for $0 \leq t_I \leq T_R$, and $\overline{h}$ and $\underline{h}$ in the proposition, we find that when $m/(c_R + r_H) \leq \underline{h}$, $\partial\pi_H/\partial t_I \leq 0$, an instant recall might still be preferred by a firm (i.e., the other situation in Case (B)). However, $m/(c_R+r_H) > \overline{h}$, $\partial\pi_H/\partial t_I > 0$ (i.e., a firm can always gain a higher profit by delaying the recall) (i.e., Case (C1)).

When $\underline{h} < m/(c_R+r_H) \leq \overline{h}$, the firm may or may not delay the recall, depending on the value of t_I. Specifically, we first find the first-order derivative of $h(t_I)$ with respect to t_I:

$$\frac{\partial h(t_I)}{\partial t_I} = \frac{f_H(t_I)}{\gamma_0(dT_R/dt_I)}\left\{[\lambda(I) - \gamma - 2\lambda(I)F_H(t_I)] - \frac{d(dT_R/dt_I)/dt_I}{(dT_R/dt_I)}\right\},$$

in which, using simple algebra, we find $[d(dT_R/dt_I)/dt_I]/(dT_R/dt_I) < 0$. When $t_I \leq T_m$, sales increase in t_I, thus leading to $f'_H(t_I) \geq 0$, which gives $\lambda(I) - \gamma - 2\lambda(I)F_H(t_I) \geq 0$, and, hence, $\partial h(t_I)/\partial t_I > 0$. That is, $h(t_I)$ is always increasing in t_I when $t_I \leq T_m$, and might be increasing

[A1]The large value of N could ensure that the Bass-diffusion model is a deterministic fluid model for which the law of large numbers applies and, hence, the probability of purchase approaches actual sales.

[A2]It could still be possible that a firm implements a delayed recall as dT_R/dt_R could become positive for some $t_R > t_I$ (i.e., for $t_R \in [t_I, T_R]$, π_H could be first decreasing, then increasing, and finally decreasing in t_R (by the assumption implying that $\partial\pi_H/\partial t_R < 0$ at $t_R = T_R$)). Therefore, the optimal recall timing that maximizes π_H could be t_I or the interior solution that satisfies $\partial\pi_H/\partial t_R = 0$ and $\partial^2\pi_H/\partial t_R^2 < 0$.

or decreasing in t_I when $t_I > T_m$.[A3] Therefore, when $\underline{h} < m/(c_R + r_H) \leq \overline{h}$, there exists at least one solution of $\tilde{t}_I$ such that $h(\tilde{t}_I) = m/(c_R + r_H)$. Thus, if $t_I < \tilde{t}_I$, a firm can always gain a higher profit by delaying its profit recall (i.e., Case (C2)).■

Proof of Corollary 1. The result for the non-delayed firm is obvious, because an instant recall is always initiated at the ending time of investigation (i.e., $t_R^* = t_I = t_C + I$ for which $I = I_0 - e$), which is increasing in t_C. However, when a firm decides to implement a delayed recall, we know $\partial\pi_H/\partial t_R > 0$ under $t_R = t_I$ (from Proposition 1) and $\partial\pi_H/\partial t_R < 0$ under $t_R = T_R$ (an assumption made earlier to avoid recalls made after the product terminates its production), implying that profit increases around t_I and decreases around T_R. Combined with the continuity of function $\partial\pi_H/\partial t_R$, we know that there exists at least one time point within $[t_I, T_R]$ that makes $\partial\pi_H/\partial t_R = 0$, as well as one that will be the optimal recall timing (denoted by t_R^*) that maximizes the value of π_H (i.e., $\partial\pi/\partial t_R^* = 0$). To analyze how t_R^* changes with t_C, we only need to observe the sign of $\partial\pi_H/(\partial t_R^* \partial t_C)$:

$$\frac{\partial^2 \pi_H}{\partial t_R^* \partial t_C} = N c_R \gamma_0 \frac{m}{c_R + r_H} \frac{\partial(dT_R/dt_R^*)}{\partial t_C} + \delta''(t_R^* - t_I), \tag{A1}$$

in which $t_I = t_C + I$, and:

$$\begin{aligned} \frac{\partial(dT_R/dt_R^*)}{\partial t_C} = {} & \frac{\lambda''(\bar{l})F_H(T_R)[1 - F_H(T_R)]f'_H(T_R)}{[f'_H(T_R)]^2} + \frac{[\gamma''(\bar{\tau}) - \lambda'(\bar{l})f_H(T_R)]f'_H(T_R)}{[f'_H(T_R)]^2} \cdot [1 - F_H(t_R)] \cdot \frac{dT_R}{dt_C} \\ & - \frac{[\gamma'(\bar{\tau}) - \lambda'(\bar{l})F_H(T_R)]\{f''_H(T_R) + f'_H(T_R)[\gamma(\bar{\tau}) + \lambda(\bar{l})F_H(T_R)]\}}{[f'_H(T_R)]^2} \cdot [1 - F_H(t_R)] \cdot \frac{dT_R}{dt_C} \end{aligned}$$

and $dT_R/dt_C = \lambda'(\bar{l})F_H(T_R)[1 - F_H(T_R)]/f'_H(T_R) > 0$. The first term on the right-hand side is greater than zero as $\lambda''(\bar{l}) \leq 0$ and $f'_H(T_R) < 0$. However, the sum of the last two terms may or may not be greater than zero. Specifically, if $[\gamma''(\bar{\tau}) - \lambda'(\bar{l})f_H(T_R)]f'_H(T_R) - [\gamma'(\bar{\tau}) - \lambda'(\bar{l})F_H(T_R)]\{f''_H(T_R) + f'_H(T_R)[\gamma(\bar{\tau}) + \lambda(\bar{l})F_H(T_R)]\} \geq 0$, it will not be lower than zero, and thus $\partial(dT_R/dt_R^*)/\partial t_C$ is always greater than zero. Thus, $\partial^2\pi_H/(\partial t_R^* \partial t_C)$ is greater than zero, implying that t_R^* increases with t_C.■

Proof of Proposition 2. Differentiating π with respect to e yields:

$$\frac{\partial \pi}{\partial e} = p\frac{\partial \pi_H(t_R^*)}{\partial e} + (1-p)\frac{\partial \pi_L}{\partial e} - c'(e),$$

in which $\partial\pi_H(t_R^*)/\partial e = [\partial\pi_H(t_R^*)/\partial t_R^*](dt_R^*/de) + \delta'(t_R^* - t_I)(\partial t_I/\partial e)$ and $\partial\pi_L/\partial e = mNf_L(T_R)(dT_R/de) + r_L N f_L(t_I)$. First, it is easy to know that $\partial\pi_L/\partial e$ is always greater than zero. Specifically, un-

[A3] A sufficient condition for $h(t_I)$ always increasing in t_I is $\lambda - \gamma - 2\lambda F_H(t_I) + \lambda(I)[1 - F_H(T_R)] > 0$.

der $\partial \pi_L/\partial e$, $dT_R/de = [1 - F_L(T_R)]\lambda'(I) F_L(T_R)/f'_L(T_R) > 0$, in which $\lambda'(I) < 0$ and $f'_L(T_R) = -f_L(T_R)[\gamma + \lambda(I) F_L(T_R)] + [1 - F_L(T_R)]\lambda(I) F_L(T_R) < 0$, and thus, $\partial \pi_L/\partial e = mN f_L(T_R)(dT_R/de) + r_L N f_L(t_I) > 0$. It then indicates that a firm facing a low-harm product defect always has incentives to exert a costly effort to speed up its investigation. Moreover, when p is smaller, e^* is higher.

Next, we study the sign of $\partial \pi_H(t_R^*)/\partial e$. From the proof of Proposition 1, we know that when a firm falls into Cases (A) or (B), we have $\partial \pi_H/\partial t_R|_{t_R^* = t_I} \leq 0$, which indicates that an instant recall ($t_R^* = t_I$) with an earlier identification time (i.e., t_I is smaller) can increase the profit π_H. As $t_I = t_C + I$ and I can be shortened by a costly effort e (i.e., $I = I_0 - e$ is decreasing in e), we know that firms in Cases (A) and (B) will exert a costly effort e to shorten investigation duration; that is, $\partial \pi_H(t_R^*)/\partial e = [\partial \pi_H(t_R^*)/\partial t_R^*](dt_R^*/de) \geq 0$, in which $\partial \pi_H(t_R^*)/\partial t_R^* \leq 0$ and $dt_R^*/de < 0$ as $t_R^* = t_I = t_C + I_0 - e$. Combined with the value of $\partial \pi_L/\partial e$ and $c'(0) = 0$, firms in Cases (A) or (B) will always exert a costly effort e to shorten investigation duration.

However, when a firm falls into Cases (C1) and (C2), we have $\partial \pi_H/\partial t_R|_{t_R = t_I} > 0$, and the optimal recall timing t_R^* should satisfy $\partial \pi_H(t_R^*)/\partial t_R^* = 0$; thus, $\partial \pi_H(t_R^*)/\partial e = -\delta'(t_R^* - t_I)$, which suppresses a firm's incentive to exert investigation efforts. In turn, the only incentive to motivate the firm to exert an investigation effort is having a high probability of a low-harm defect (i.e., $(1-p)\partial \pi_L/\partial e > 0$). Compared with Cases (A) and (B), a firm falling into Cases (C1) and (C2) always has a lower incentive to shorten investigation duration than firms in Cases (A) or (B), given the same defect notice time t_C, the same margin m, and the same low-harm compensation r_L. In particular, if $p = 1$, i.e., then a high-harm risk of defect is always realized, and the firm in Cases (C1) and (C2) will not exert any costly effort to speed up an investigation (i.e., $e^* = 0$).■

Proof of Corollary 2. To begin, we examine the sign of $\partial^2 \pi/(\partial e \partial t_C)$, which is:

$$\frac{\partial^2 \pi}{\partial e \partial t_C} = p\frac{\partial^2 \pi_H(t_R^*)}{\partial e \partial t_C} + (1-p)mN\left[f'_L(T_R)\frac{dT_R}{dt_C}\frac{dT_R}{de} + f_L(T_R)\frac{\partial(dT_R/de)}{\partial t_C}\right] + (1-p)r_L N f'_L(t_I)\frac{\partial t_I}{\partial t_C},$$

in which $t_I = t_C + I_0 - e$, $\partial t_I/\partial t_C > 0$, and $f'_L(T_R)(dT_R/dt_C)(dT_R/de) + f_L(T_R)[\partial(dT_R/de)]/\partial t_C < 0$, which can be proved with simple algebra. From the proof of Corollary 1, we know that for a delayed firm whose recall timing increases with the defect notice time, the first term, $p\partial^2 \pi_H(t_R^*)/(\partial e \partial t_C) = -\delta''(t_R^* - t_I)[\partial t_R^*/\partial t_C - \partial t_I/\partial t_C] < 0$. However, the sign of the last term, $r_L N f'_L(t_I)\partial t_I/\partial t_C$, is ambiguous, as $\partial t_I/\partial t_C > 0$ and $f'_L(t_I)$ depends on t_I. Clearly, when $t_I > T_m$, $f'_L(t_I) < 0$, and thus, $\partial^2 \pi/(\partial e \partial t_C) < 0$. That is, e^* decreases with t_C when $t_I > T_m$.

For a non-delayed firm (in which $t_R^* = t_I$), we again examine $\partial^2\pi/(\partial e \partial t_C)$. We note that:

$$\frac{\partial^2 \pi_H(t_R^*)}{\partial e \partial t_C} = \frac{\partial^2 \pi_H}{\partial t_I^2}\frac{\partial t_I}{\partial t_C}\frac{\partial t_I}{\partial e} + \frac{\partial \pi_H}{\partial t_I}\frac{\partial^2 t_I}{\partial e \partial t_C} = -\frac{\partial t_I}{\partial t_C} N[m\gamma_0 \frac{\partial (dT_R/dt_I)}{\partial t_I} - (c_R + r_H) f_H'(t_I)],$$

in which $\partial^2 t_I/(\partial e \partial t_C) = 0$, $\partial t_I/\partial t_C > 0$, $\partial(dT_R/dt_I)/\partial t_I > 0$ when $\gamma'(\overline{\tau}) < \lambda'(I)F_H(T_R)$ (from the proof of Proposition 1)[A4], and $f_H'(t_I) < 0$ when $t_I > T_m$. When combined with our preceding analysis for a delayed firm, we know that a non-delayed firm with $\gamma'(\overline{\tau}) < \lambda'(I)F_H(T_R)$ always exerts a higher investigation effort if a defect is noticed at an earlier time but is identified after the mature time.■

Proof of Proposition 3. The first part of this proof is the same as the proof of Corollary 1 and Corollary 2 because $\partial t_I/\partial t_C > 0$ in these cases.

Next, we prove the second part. First, we prove the optimal recall timing t_R^*, given an investigation effort e. For a non-delayed firm, as it initiates an prompt recall (i.e., $t_R^* = t_I = t_C + I_0(t_C) - e$), an earlier defect notice time then leads to a later recall time as $\partial t_I/\partial t_C < 0$. For a delayed firm, we observe the sign of $\partial \pi_H/(\partial t_R^* \partial t_C)$, which is:

$$\frac{\partial^2 \pi_H}{\partial t_R^* \partial t_C} = N c_R \gamma_0 \frac{m}{c_R + r_H}\frac{\partial (dT_R/dt_R^*)}{\partial t_C} + \delta''(t_R^* - t_I) + \delta''(t_R^* - t_I) I_0'(t_C), \tag{A2}$$

in which $t_I = t_C + I_0(t_C) - e$. We note that Equation (A2) (with a learning effect) differs from Equation (A1) (without a learning effect) only in $\delta''(t_R^* - t_I)I_0'(t_C)$, in which the learning effect during an investigation provides a negative incentive for accelerating the recall timing when t_C occurs earlier. That is, the value $\partial^2\pi_H/(\partial t_R^* \partial t_C)$ can be below zero when $\delta''(t_R^* - t_I)I_0'(t_C)$ is sufficiently small (i.e., t_R^* may decrease with t_C for some t_C). Thus, the optimal recall timing t_R^* may or may not be earlier if t_C is earlier.

Next, to study the investigation effort, we must study the sign of $\partial^2\pi/(\partial e \partial t_C)$:

$$\frac{\partial^2 \pi}{\partial e \partial t_C} = p\frac{\partial^2 \pi_H(t_R^*)}{\partial e \partial t_C} + (1-p)mN\left[f_L'(T_R)\frac{dT_R}{dt_C}\frac{dT_R}{de} + f_L(T_R)\frac{\partial(dT_R/de)}{\partial t_C}\right] + (1-p)r_L N f_L'(t_I)\frac{\partial t_I}{\partial t_C},$$

in which $t_I = t_C + I_0(t_C) - e$, $\partial t_I/\partial t_C < 0$, $\partial^2\pi_H(t_R^*)/(\partial e \partial t_C) < 0$, and $f_L'(T_R)(dT_R/dt_C)(dT_R/de) + f_L(T_R)[\partial(dT_R/de)]/\partial t_C < 0$, which can be proved with simple algebra. Clearly, when $t_I < T_m$, $f_L'(t_I) > 0$, and thus, $\partial^2\pi/(\partial e \partial t_C) < 0$. That is, the optimal investigation effort e^* decreases with t_C when $t_I < T_m$. However, when $t_I > T_m$, $f_L'(t_I) < 0$, and thus, $\partial^2\pi/(\partial e \partial t_C)$ can be greater than zero for some high values of $r_L N f_L'(t_I)(\partial t_I/\partial t_C)$ (i.e., the firm may or may not exert a higher effort

[A4] When $\gamma'(\overline{\tau}) \geq \lambda'(I)F_H(T_R)$, we can not always have $\partial(dT_R/dt_T)/\partial t_I > 0$.

for an earlier defect notice time when $t_I > T_m$).

For a non-delayed firm (in which $t_R^* = t_I$), we again examine the sign of $\partial^2\pi/(\partial e\partial t_C)$:

$$\frac{\partial^2\pi_H(t_R^*)}{\partial e\partial t_C} = \frac{\partial^2\pi_H}{\partial t_I^2}\frac{\partial t_I}{\partial t_C}\frac{\partial t_I}{\partial e} + \frac{\partial\pi_H}{\partial t_I}\frac{\partial^2 t_I}{\partial e\partial t_C} = -\frac{\partial t_I}{\partial t_C}N[m\gamma_0\frac{\partial(dT_R/dt_I)}{\partial t_I} - (c_R + r_H)f_H'(t_I)],$$

in which $\partial^2 t_I/(\partial e\partial t_C) = 0$, $\partial t_I/\partial t_C < 0$, $\partial(dT_R/dt_I)/\partial t_I > 0$ when $\gamma'(\overline{\tau}) < \lambda'(I)F_H(T_R)$, and $f_H'(t_I) < 0$ when $t_I > T_m$. That is, $\partial^2\pi_H(t_R^*)/(\partial e\partial t_C) > 0$ when $t_I > T_m$. When combined with the preceding analysis for a delayed firm (in which case the sign of $\partial^2\pi_L/(\partial e\partial t_C)$ remain ambiguous when $t_I > T_m$), we know that a non-delayed firm with $\gamma'(\overline{\tau}) < \lambda'(I)F_H(T_R)$ may exert a lower investigation effort if a defect is noticed at an earlier time but is identified after the mature time.■

Proof of Proposition 4. When $\rho = 0$, the signal does not provide additional information, and hence, we have $e^{**} = e^*$. When the signal provides some information ($0 < \rho \leq 1$), the signal can help the firm improve its understanding of the defect type. The time of receiving a signal ($\hat{t}$) is assumed to be close to the defect notice time, which ensures sufficient time for the firm to make effort adjustments. When the firm receives a high-harm signal ($\sigma = \tilde{H}$), the posterior probability of the high-harm defect is $Pr(H|\tilde{H}) = \rho + (1-\rho)Pr(H) = p + \rho(1-p) > p$. As implied by Proposition 2, a higher probability of having a high-harm defect makes the delayed firm exert less effort in an investigation compared with the case without a signal (i.e., $e^{**} < e^*$). When receiving a low-harm signal ($\sigma = \tilde{L}$), $Pr(H|\tilde{L}) = (1-\rho)Pr(H) < p$ leads to $e^{**} > e^*$. Moreover, when $\rho = 1$, a high-harm signal implies that the posterior probability of the high-harm status is one, and thus the effort for the delayed firm becomes zero, as implied by Proposition 2.

For the non-delayed firm, the effort comparison depends on the marginal benefits of investigation effort for high-harm and low-harm profits. Specifically, differentiating π with respect to e yields: $\partial\pi/\partial e = \partial\hat{\pi}_L/\partial e + Pr(H|\sigma)(\partial\hat{\pi}_H/\partial e - \partial\hat{\pi}_L/\partial e) - c'(e)$. We note that $\partial\hat{\pi}_H/\partial e = \partial\pi_H/\partial e$ and $\partial\hat{\pi}_L/\partial e = \partial\pi_H/\partial e$ as the exogenous signal arrival time, $\hat{t}$, does not affect the firm's incentive. Thus, if the marginal benefit of investigation is higher for a high-harm defect (i.e., $\partial\hat{\pi}_H/\partial e > \partial\hat{\pi}_L/\partial e$), then a high-harm signal leads the non-delayed firm to exert more effort with respect to an investigation compared with the situation without a signal (i.e., $Pr(H|\sigma) > p$ and thus $e^{**} > e^*$). However, a low-harm signal makes it exert less effort with respect to an investigation as the probability of a low-harm signal is higher now (i.e., $Pr(H|\sigma) < p$ and thus $e^{**} < e^*$). However, when $\partial\hat{\pi}_H/\partial e \leq \partial\hat{\pi}_L/\partial e$, the non-delayed firm will behave the same as the delayed firm (i.e., exerting more effort ($e^{**} > e^*$) when receiving a low-harm signal, but exerting less effort ($e^{**} < e^*$) when receiving a high-harm signal). ■

Proof of Corollary 3. Consider two penalty functions, $\delta_L(\cdot)$, $\delta_H(\cdot)$, in which $\delta'_L(t_R - t_I) < \delta'_H(t_R - t_I)$ for the same delay duration $t_R - t_I > 0$. We denote t_L^* (t_H^*) as the optimal recall timing for $\delta'_L(\cdot)$ ($\delta'_H(\cdot)$). By definition, we have $\partial\pi/\partial t_L^* = N[mf_H(T_R)dT_R/dt_L^* - (c_R + r_H)f_H(t_L^*)] - \delta'_L(t_L^* - t_I) = 0$. By replacing $\delta'_L(t_L^* - t_I)$ with $\delta'_H(t_L^* - t_I)$, we have $\partial\pi_H/\partial t_L^* = N[mf_H(T_R)dT_R/dt_L^* - (c_R + r_H)f_H(t_L^*)] - \delta'_H(t_L^* - t_I)$. We note that the optimal recall timing should satisfy the second-order derivative condition: $\partial^2\pi_H/\partial(t_R^*)^2 < 0$. It then implies that $t_H^* < t_L^*$. ■

Proof of Proposition 5. The analysis for a lump-sum penalty B can be found in Appendix B. First, we let π_I be the profit function with a penalty function of $\delta(t_R - t_I, \Delta I)$, given that a high-harm defect is realized. As only the penalty term in π_I involves ΔI, to find the optimal manipulated investigation duration ΔI^*, we first show that the condition in Proposition 5 leads to the concavity of π_I, and then find the maximum of π_I by varying ΔI. The first-order and second-order derivatives of $\delta(t_R - t_I, \Delta I)$ are $d\pi_I/d\Delta I = -\partial\delta(t_R - t_I, \Delta I)/\partial\Delta I + \partial\delta(t_R - t_I, \Delta I)/\partial(t_R - t_I)$ and $d^2\pi_I/d\Delta I^2 = -\partial^2\delta(t_R - t_I, \Delta I)/\partial\Delta I^2 + \partial^2\delta(t_R - t_I, \Delta I)/[\partial(t_R - t_I)\partial\Delta I] - \partial^2\delta(t_R - t_I, \Delta I)/\partial(t_R - t_I)^2 < 0$, in which $t_I = t_C + \tilde{I} = t_C + I + \Delta I$. Therefore, ΔI^* satisfies that $d\pi_I/d\Delta I^* = 0$. Compared to the penalty without ΔI, in which the first-order condition always leads to as high as possible ΔI, ΔI^* is always smaller with the penalty with ΔI.

Given ΔI^*, the firm decides its (delayed) recall timing (denoted by t_R^*, and $t_R^* > t_C + I$) by using the first-order condition: $\partial\pi_I/\partial t_R^* = N[mf_H(T_R)dT_R/dt_R^* - (c_R + r_H)f_H(t_R^*)] - \partial\delta(t_R^* - t_I, \Delta I)/\partial(t_R^* - t_I) = 0$. Therefore, we know the first-order condition of the profit function π_H with $\delta(t_R - t_I)$ is:

$$\frac{\partial\pi_H}{\partial t_R^*} = N[mf_H(T_R)\frac{dT_R}{dt_R^*} - (c_R + r_H)f_H(t_R^*)] - \frac{\partial\delta(t_R^* - t_I)}{\partial(t_R^* - t_I)} = \frac{\partial\delta(t_R^* - t_I, \Delta I)}{\partial(t_R^* - t_I)} - \frac{\partial\delta(t_R^* - t_I)}{\partial(t_R^* - t_I)} > 0,$$

in which the inequality follows the assumption that $\partial^2\delta/\partial(t_R - t_I)\partial\Delta I > 0$ and $\delta(t_R - t_I, 0) = \delta(t_R - t_I)$. ■

Appendix B: Analysis for Penalties

In Section 4.3, we discuss the case in which a delayed firm can manipulate its investigation duration to avoid a heavy government penalty. To reduce such opportunistic behavior, we consider two penalty schemes: penalizing based on the manipulated investigation duration and based on manipulative behaviors. In the following sections, we will show the effectiveness of the two schemes.

B.1 Penalty Based on Manipulated Investigation Duration

We assume that the penalty function takes a form of $\delta(t_R - t_I, \Delta I)$, in which $t_I = t_C + \tilde{I}$ and $\Delta I = \tilde{I} - I$. It follows similar assumptions as the one without ΔI (i.e., $\delta(0, \Delta I) > 0$, $\partial\delta/\partial\Delta I > 0$, and $\partial^2\delta/\partial\Delta I^2 > 0$), and we also assume $\delta(t_R - t_I, 0) = \delta(t_R - t_I) > 0$ to be consistent. When a regulator can only penalize a firm's misconduct for its recall delays (i.e., $t_R - t_I$), a longer manipulated investigation duration always leads to a further delayed recall. However, when a regulator can penalize a firm for both recall delays as well as manipulated investigation duration, a firm can no longer pretend to conduct a long investigation solely to avoid the heavy penalty from long recall delays. Depending on the interactions between the delay duration, $t_R - t_I$, and the manipulated investigation duration ΔI, firms may react differently. In Proposition 5, we have shown sufficient conditions under which incorporating manipulated investigation duration into the penalty function reduces recall delays.

Recall that firms are motivated to manipulate investigations in order to reduce penalties related to delay duration, and they determine their recall timings by trading off the marginal benefit of accumulated sales from a delayed recall and the marginal cost of recall costs and government penalties.[A5] The condition implies that not only do the recall delay and the manipulated investigation duration need to be complements for raising penalties (i.e., $\partial^2\delta/[\partial(t_R - t_I)\partial\Delta I] > 0$), but that the complementarity cannot be too high so as to ensure an effective penalty scheme for deterring both recall delays and manipulated investigation duration. Otherwise, if $\partial^2\delta/[\partial(t_R - t_I)\partial\Delta I] \geq \partial^2\delta/[\partial(t_R - t_I)]^2 + \partial^2\delta/[\partial\Delta I]^2$, the firm may choose to either push the manipulated investigation duration to its limits and enjoy a low delay duration or delay a recall as much as possible, but nonetheless report truthfully to reduce manipulation penalties.

B.2 Penalty Based on Manipulation Behaviors

Although imposing penalties directly on firms for their misconduct effectively reduces manipulation and recall delays, it is highly costly to pin down the true investigation duration; moreover, doing so may not completely deter manipulations. Next, we consider a lump-sum penalty (defined by $B > 0$) based on whether a firm commits manipulation, which can largely relieve a regulator from costly investigations with respect to exact identification times. It only requires that the regulator

[A5] The recall timing is given by the first-order condition: $\partial\pi_H/\partial t_R = N[mf_H(T_R)dT_R/dt_R - c_R f_H(t_R)] - \partial[\delta(t_R - t_I, \Delta I)]/\partial t_R$. The marginal benefit ($Nmf_H(T_R)dT_R/dt_R$) is determined by t_R, whereas the marginal cost ($c_R f_H(t_R) + \partial[\delta(t_R - t_I, \Delta I)]/\partial t_R$) is jointly determined by t_R and ΔI.

find evidence to verify a firm's fraudulent behavior, such as concealing vital information/reports about product defects and secretly changing designs or repairing defects. Using evidence, regulators can even file legal proceedings against a firm for intentional concealment or misleading consumers/regulators. Specifically, we assume that the probability of verifying manipulation (denoted by $q(\Delta I)$) increases with the manipulated investigation duration ΔI, as it is easy to speculate that the firm is more likely to be exposed under a longer manipulated investigation duration.

Next, we denote $\pi_H^M = \max_{t_R, \Delta I} \; Nm[F_H(T_R) - F_H(t_C)] - (c_R + r_H)F_H(t_R) - \delta(t_R - t_C - I - \Delta I) - q(\Delta I)B$, which is the firm's maximum profit, by manipulating both the investigation duration and recall timing. We can see from the cross-derivative, $\partial^2 \pi_H^M / (\partial \Delta I \partial B) = -\partial q(\Delta I)/\partial \Delta I < 0$, that the marginal benefit of manipulating a longer investigation duration (i.e., $\partial \pi_H^M / \partial \Delta I$) is decreasing in the lump-sum penalty B, implying that a higher B leads to a shorter investigation manipulation. Moreover, if we compare the maximum profit involving manipulation (i.e., π_H^M) with that involving no manipulation (i.e., $\pi_H^N = \max_{t_R} \; Nm[F_H(T_R) - F_H(t_C)] - (c_R + r_H)F_H(t_R) - \delta(t_R - t_C - I)$), we find that if B is sufficiently large, the firm will have $\pi_H^M < \pi_H^N$, implying that the firm will not commit any investigation duration manipulation. However, we note that this condition of B is case-dependent, as it involves understanding a firm's cost structure and may subject the firm to enforcement difficulty if the fine is too high. Moreover, regulators often need to set rules prior to any verified misconduct. As a result, for some cases (e.g., firms with a lower profit gain from manipulation), a pre-set penalty B may completely eliminate manipulation, whereas for others (e.g., firms with a greater gain from manipulation), it only partially helps reduce these opportunistic behaviors.

Our result helps explain what happened after Toyota's 2009-2010 recall. In addition to paying \$17.35 million for the delayed recall, Toyota paid \$1.2 billion to settle a criminal charge in 2014 for hiding safety defects from the public. After the heavy lump-sum fines on Toyota's recall, auto makers such as Ford, Chrysler, Honda, and Nissan not only sped up their own investigations, but also became more willing to issue recalls to avoid risks of massive fines, criminal charges, and a rash of lawsuits, which would have made "2014 a record year for recalls."[A6] In summary, a long product cycle offers firms multiple chances to take opportunistic action. Therefore, when designing penalty schemes for delayed recalls, regulators must take these opportunistic behaviors into account.

[A6] https://www.latimes.com/business/autos/la-fi-hy-recall-fever-20140516-story.html