# Online Appendix: Regulating Transformative Technologies

Daron Acemoglu    Todd Lensman

July 8, 2024

Appendix A contains proofs for the results in the main text. In Appendices B, C, and D, we analyze extensions of our benchmark model and discuss the robustness of our main results.

# A Proofs for the Main Text

In this part of the Appendix, we provide proofs for results in the main text.

## A.1 Proofs for Section II

**Proof of Proposition 2.** Let $L(t) \equiv L(\mu(t), q(t))$ denote the damage threshold at time $t \in (0, \infty)$. Making use of Proposition 1 and the equation $q(t) = q(0) + (g_N - g_O)t$, the damage threshold equals

$$L(t) = \alpha + (\rho - g_N)\left[\frac{\alpha - \exp(-[q(0) + (g_N - g_O)t])}{\mu(t)\lambda\eta} - \frac{\exp(-[q(0) + (g_N - g_O)t])}{\rho - g_O}\right]. \tag{A1}$$

It is immediate that $L(t)$ is strictly decreasing in $g_O$, so adoption at time $t$ is nonincreasing in $g_O$. Note that adoption is also strictly decreasing whenever $L(t) \in \left(\underline{\delta}, \bar{\delta}\right)$.

Considering instead the comparative static with respect to $g_N$, we can differentiate to find

$$\frac{\partial L(t)}{\partial g_N} = (1 + (\rho - g_N)t)\left[\frac{1}{\mu(t)\lambda\eta} + \frac{1}{\rho - g_O}\right]\exp(-[q(0) + (g_N - g_O)t]) - \frac{\alpha}{\mu(t)\lambda\eta}.$$

This derivative is positive iff

$$(1 + (\rho - g_N)t)\left[1 + \frac{\mu(t)\lambda\eta}{\rho - g_O}\right]\exp(-[z(0) + (g_N - g_O)t]) \geq \alpha.$$

The left side limits to zero as $t \to \infty$, so there exists an earliest time $\bar{t} < \infty$ such that $\partial L(t)/\partial g_N < 0$ for $t > \bar{t}$. If $\bar{t} > 0$, then the left side of the inequality above must be decreasing in $t$ at $t = \bar{t}$. Since the left side is also decreasing in $g_N$, we must have that $\bar{t}$ is decreasing in $g_N$.

Finally, the bracketed term in (A1) limits to a finite value as $g_N$ increases to $\rho$, which implies that $L(t)$ limits to $\alpha$. Since the lower support of $F$ satisfies $\alpha \leq \underline{\delta}$, we conclude that $X(\mu(t), q(t))$ limits to zero. ■

**Proof of Proposition 3.** Using the expression for the damage threshold (8), we can calculate

$$\frac{\dot{L}(\mu,q)}{\rho-g_N} = \frac{1-\mu}{\mu}\frac{\alpha-\exp(-q)}{\eta} + \left(\frac{1}{\mu\lambda\eta} + \frac{1}{\rho-g_O}\right)(g_N-g_O)\exp(-q).$$

This equation implies that $\dot{L}(\mu,q)$ is strictly decreasing in $g_O$. Differentiating implies that $\dot{L}(\mu,q)$ is strictly decreasing in $g_N$ iff

$$\alpha\exp(q) - 1 > \frac{(\rho-g_N)-(g_N-g_O)}{1-\mu}\left(\frac{1}{\lambda} + \frac{\mu\eta}{\rho-g_O}\right).$$

Differentiating $\dot{L}(\mu,q)$ again yields

$$\frac{\ddot{L}(\mu,q)}{\rho-g_N} = \lambda\frac{1-\mu}{\mu}\frac{\alpha-\exp(-q)}{\eta} + \left[2\frac{1-\mu}{\mu} - \left(\frac{1}{\mu\lambda} + \frac{\eta}{\rho-g_O}\right)(g_N-g_O)\right](g_N-g_O)\frac{\exp(-q)}{\eta}.$$

Provided that $\alpha > \exp(-q)$, this expression immediately implies that $\ddot{L}(\mu,q) < 0$ iff $g_N - g_O > G(\mu,q)$, where $G(\mu,q)$ is the largest solution to the quadratic equation

$$\lambda\frac{1-\mu}{\mu}\frac{\alpha-\exp(-q)}{\eta} + \left[2\frac{1-\mu}{\mu} - \left(\frac{1}{\mu\lambda} + \frac{\eta}{\rho-g_O}\right)G(\mu,q)\right]G(\mu,q)\frac{\exp(-q)}{\eta} = 0.$$

Equivalently,

$$G(\mu,q) = \lambda\frac{1+\sqrt{1+\left(1+\frac{\mu\lambda\eta}{\rho-g_O}\right)(1-\mu)^{-1}(\alpha\exp(q)-1)}}{\left(1+\frac{\mu\lambda\eta}{\rho-g_O}\right)(1-\mu)^{-1}}.$$

We observe that $G$ is decreasing in $\mu$ and increasing in $q$, so that $\dot{G}(\mu,q) > 0$ with

$$\lim_{t\to\infty} G(\mu(t), q(t)) = \infty.$$

■

**Proof of Proposition 4.** See the proof of Proposition C.2 in Appendix C. ■

## A.2 Proofs for Section IV

**Proof of Proposition 8.** With a sector-independent use tax $\tau(\mu, Q)$, it is privately optimal to use technology $N$ before the disaster iff

$$\alpha Q_N - Q_O - \tau(\mu, Q) > \mu\lambda\eta\left[\frac{1}{\rho - g_O}Q_O - \frac{\alpha - \gamma_i}{\rho - g_N}Q_N\right]. \tag{A2}$$

The right side of this inequality is strictly increasing in $\gamma_i$. Given an initial state $(\mu(0), Q(0))$, let $\tilde{t}_i$ denote the time at which firm $i$ begins using technology $N$. For any sector $j$ with private damages $\gamma_j$, we immediately observe that $\gamma_i \leq \gamma_j$ iff $\tilde{t}_i \leq \tilde{t}_j$. The latter inequality is strict if $\gamma_i < \gamma_j$ and $\tilde{t}_j > 0$.

If $\gamma_i$ and $\delta_i$ are positively affiliated, the tax (13) suffices to implement socially optimal technology choices in equilibrium. To see this, note that the private optimality condition (A2) implies that firm $i$ uses technology $N$ in state $(\mu, Q)$ iff $\gamma_i < \hat{L}(\mu, q)$, where

$$\frac{\hat{L}(\mu, q) - \alpha + L(\mu, q) - \kappa(L(\mu, q))}{\rho - g_N} = \frac{\alpha - \exp(-q)}{\mu\lambda\eta} - \frac{\exp(-q)}{\rho - g_O}.$$

Using the definition of the damage threshold $L(\mu, q)$ from Proposition 1, this equation reduces to $\hat{L}(\mu, q) = \kappa(L(\mu, q))$. Since $\kappa$ is strictly increasing, we conclude that equilibrium technology choices are efficient: Firm $i$ uses technology $N$ iff

$$\delta_i = \kappa^{-1}(\gamma_i) < \kappa^{-1}\left(\hat{L}(\mu, q)\right) = L(\mu, q).$$

Finally, fix an initial state $(\mu(0), Q(0))$ such that $L(\mu(0), Q(0)) < \underline{\delta}$, so that it is inefficient for any sector to use technology $N$ at $t = 0$. Suppose that a given sector-independent tax $\tau(\mu, Q)$ implements socially optimal technology choices in equilibrium. We can define the affiliation function $\kappa$ as follows: For any value of social damages $\delta \in \left[\underline{\delta}, \bar{\delta}\right]$, let $t(\delta) > 0$ be the time at which sectors with social damages $\delta$ (socially) optimally begin using technology $N$. Since $\tau(\mu, Q)$ implements socially optimal technology choices in equilibrium, these same sectors must find it privately optimal to begin using technology $N$ at time $t(\delta)$. These sectors must have a common value of private damages $\gamma(t(\delta))$: If one sector had a larger value of private damages $\gamma' > \gamma(t(\delta))$, it would find it privately optimal to delay using technology $N$, contradicting the assumption that $\tau$ implements socially optimal technology choices. As a result, the affiliation function $\kappa(\delta) = \gamma(t(\delta))$ is well-defined, and we conclude that social and private damages must be positively affiliated. ■

**Proof of Proposition 9.** Given a threshold $\hat{\delta}$ and wait time $\hat{T}$, the planner's objective discounted to $t = 0$ can be written

$$
\begin{aligned}
V\left(\hat{\delta},\hat{T}\right) = \int_0^{\hat{T}} \exp\left(-\rho t\right) \int_{\delta_i<\hat{\delta}} &\Big\{\left(1 - x\left(\mu(t), q(t), \gamma_i\right)\right)\left[1 + \mu(t)\lambda\eta\frac{1}{\rho - g_O}\right]Q_O(t) \\
&+ x\left(\mu(t), q(t), \gamma_i\right)\left[\alpha + \mu(t)\lambda\eta\frac{\alpha - \delta_i}{\rho - g_N}\right]Q_N(t)\Big\}\, di dt \\
+ \int_0^{\hat{T}} \exp\left(-\rho t\right) \int_{\delta_i\geq\hat{\delta}} &\left[1 + \mu(t)\lambda\eta\frac{1}{\rho - g_O}\right]Q_O(t) di dt \\
+ \int_{\hat{T}}^{\infty} \exp\left(-\rho t\right) \int_0^1 &\Big\{\left(1 - x\left(\mu(t), q(t), \gamma_i\right)\right)\left[1 + \mu(t)\lambda\eta\frac{1}{\rho - g_O}\right]Q_O(t) \\
&+ x\left(\mu(t), q(t), \gamma_i\right)\left[\alpha + \mu(t)\lambda\eta\frac{\alpha - \delta_i}{\rho - g_N}\right]Q_N(t)\Big\}\, di dt.
\end{aligned}
$$

Here $x\left(\mu, q, \gamma_i\right)$ denotes the unrestricted equilibrium technology choice given state $(\mu, q)$ and private damages $\gamma_i$:

$$
x\left(\mu, q, \gamma_i\right) = \begin{cases} 1 & \text{if } \alpha_i - \exp(-q) > \mu\lambda\eta\left[\frac{1}{\rho - g_O}\exp(-q) - \frac{\alpha - \gamma_i}{\rho - g_N}\right], \\ 0 & \text{else.} \end{cases}
$$

With $\hat{\delta}$ fixed, we can differentiate $V$ with respect to $\hat{T}$ to find

$$
\exp\left(\rho\hat{T}\right)\frac{\partial V\left(\hat{\delta},\hat{T}\right)}{\partial\hat{T}} = -\int_{\delta_i\geq\hat{\delta}} x\left(\mu, q, \gamma_i\right)\left\{\alpha Q_N - Q_O - \mu\lambda\eta\left[\frac{1}{\rho - g_O}Q_O - \frac{\alpha - \delta_i}{\rho - g_N}Q_N\right]\right\} di.
$$

To simplify notation, we have left the dependence of the state $(\mu, Q)$ on the wait time $\hat{T}$ implicit. First observe that the optimal wait time is bounded:

$$
\lim_{\hat{T}\to\infty}\frac{1}{Q_N\left(\hat{T}\right)}\exp\left(\rho\hat{T}\right)\frac{\partial V\left(\hat{\delta},\hat{T}\right)}{\partial\hat{T}} = -\alpha\int_{\delta_i\geq\hat{\delta}} di < 0.
$$

Let $\tilde{t}_i$ denote the equilibrium time of adoption for sector $i$ when unrestricted, and let $\underline{t}(\hat{\delta}) \geq 0$ denote the greatest lower bound for these times across all sectors above the threshold ($\delta_i \geq \hat{\delta}$). Note that we can write

$$
\exp\left(\rho\hat{T}\right)\frac{\partial V\left(\hat{\delta},\hat{T}\right)}{\partial\hat{T}} = -\int_{\delta_i\geq\hat{\delta},\tilde{t}_i\leq\hat{T}} \alpha Q_N - Q_O - \mu\lambda\eta\left[\frac{1}{\rho - g_O}Q_O - \frac{\alpha - \delta_i}{\rho - g_N}Q_N\right] di.
$$

Clearly $\partial V(\hat{\delta},\hat{T})/\partial\hat{T} = 0$ for $\hat{T} \leq \underline{t}\left(\hat{\delta}\right)$. But $\partial V(\hat{\delta},\hat{T})/\partial\hat{T} > 0$ for $\hat{T}$ just above $\underline{t}\left(\hat{\delta}\right)$, because

$$\frac{\partial}{\partial \hat{T}} \exp\left(\rho \hat{T}\right) \frac{\partial V\left(\hat{\delta},\hat{T}\right)}{\partial \hat{T}}\Bigg|_{\hat{T}=\underline{t}(\hat{\delta})} = -\int_{\delta_i \geq \hat{\delta}, \tilde{t}_i = \underline{t}(\hat{\delta})} \alpha Q_N - Q_O - \mu\lambda\eta \left[\frac{1}{\rho - g_O} Q_O - \frac{\alpha - \delta_i}{\rho - g_N} Q_N\right] di.$$

On the right-hand side, the state $(\mu, Q)$ is evaluated at $\underline{t}(\hat{\delta})$. Since $\gamma_i < \delta_i$ for all sectors above the threshold, the right-hand side must be strictly positive. This implies that $V$ is strictly increasing in $\hat{T}$ just above $\underline{t}(\hat{\delta})$, so the optimal wait time $\hat{T}$ must be interior. It satisfies the first-order condition

$$0 = -\int_{\delta_i \geq \hat{\delta}} x\left(\mu, q, \gamma_i\right) \left\{\alpha Q_N - Q_O - \mu\lambda\eta \left[\frac{1}{\rho - g_O} Q_O - \frac{\alpha - \delta_i}{\rho - g_N} Q_N\right]\right\} di.$$

Setting $\hat{T} = \underline{t}(\hat{\delta})$ replicates the laissez-faire equilibrium, so this argument establishes that a sandbox policy with $\hat{\delta} > \underline{\delta}$ can strictly improve upon the laissez-faire equilibrium. ■

# B Extensions

In this part of the Appendix, we discuss one generalization and two extensions.

## B.1 Unrestricted Heterogeneity across Sectors

In this section, we briefly describe how our analysis generalizes without the simplifying assumption that $\alpha_i$ and $\eta_i$ are constant across sectors (but maintaining the "large damages" assumption (6)). The basic optimality condition (7) describes the solution to the planner's problem, and we can rearrange this condition to observe that there exists a *sector-specific* damage threshold $L_i(\mu,q)$ such that it is optimal to adopt the new technology in sector $i$ iff $\delta_i < L_i(\mu,q)$. This damage threshold satisfies the analogue to (8) with $\alpha_i$ and $\eta_i$ in place of $\alpha$ and $\eta$:

$$\frac{L_i(\mu,q)-\alpha_i}{\rho-g_N}=\frac{\alpha_i-\exp(-q)}{\mu\lambda\eta_i}-\frac{\exp(-q)}{\rho-g_O}.$$

The comparative statics of Proposition 1 naturally hold for this sector-specific damage threshold, and they should now be interpreted as comparative statics for the adoption of the new technology *within sector i* as opposed to economy-wide.

Adoption of the new technology is defined exactly as in the benchmark model:

$$X(\mu,q)=\int_0^1 x_i(\mu,q)di. \tag{B1}$$

It is straightforward to observe that the comparative statics of Proposition 2 apply immediately to sector-level adoption $x_i(\mu,q)=\mathbb{1}[\delta_i<L_i(\mu,q)]$ and, provided that the joint distribution of the sector characteristics $(\delta_i,\alpha_i,\eta_i)$ is well-behaved (e.g., compact support), to overall adoption $X$. The slope and curvature results of Proposition 3 extend immediately to the sector-specific damage threshold $L_i(\mu,q)$. In particular, the second part of this proposition describes a sense in which adoption is convex *within* each sector, and with sufficient dispersion in damages $\delta_i$ this provides a force for convex adoption *across* sectors. If we additionally relax the assumption that damages are large (6), the key implication of Proposition 4 holds: With $q(0)$ sufficiently low, it remains optimal to delay adoption in sector $i$ even if $g_N$ increases toward the discount rate $\rho$. The optimality of gradual adoption is robust to all of these generalizations.

Finally, we note that the same adjustments can be made to our analysis of equilibrium adoption, and similarly for our regulation results: Proposition 8 generalizes in the sense that sector-independent use taxes can restore socially optimal adoption if and only if the laissez-faire and socially optimal orders of adoption coincide. Positive affilitation of damages again suffices for this, but with heterogeneity in $\alpha_i$ and $\eta_i$, it is typically stricter than necessary. By a similar argument as in the proof of Proposition 9, it can again be verified that a regulatory sandbox can generally improve upon the laissez-faire equilibrium.

## B.2 Heterogeneous $\alpha_i$

Suppose that $\eta_i$ and $\delta_i$ are constant across sectors, and let $F_\alpha$ denote the smooth distribution function for $\alpha_i$ with support $\left[\underline{\alpha}, \bar{\alpha}\right]$. We maintain the assumption that $\alpha_i \leq \delta$ for each sector $i$, which requires $\bar{\alpha} \leq \delta$. Making use of the planner's optimality condition (7), we observe that there exists a *productivity threshold* $A(\mu, q)$ such that it is optimal to use the new technology in sector $i$ iff $\alpha_i > A(\mu, q)$. Total adoption of the new technology is then the fraction of sectors above the productivity threshold:

$$X(\mu, q) = 1 - F_\alpha(A(\mu, q)).$$

The following proposition characterizes the productivity threshold and is analogous to Proposition 1 in Section II.B.

**Proposition B.1.** *It is socially optimal to use technology $N$ in sector $i$ iff $\alpha_i > A(\mu, q)$, where*

$$A(\mu, q) + \mu\lambda\eta \frac{A(\mu, q) - \delta}{\rho - g_N} = \left(1 + \frac{\mu\lambda\eta}{\rho - g_O}\right)\exp(-q). \tag{B2}$$

*$A(\mu, q)$ (and thus $1 - X(\mu, q)$) is strictly decreasing in $q$; strictly increasing in $g_O$ and $\delta$; and strictly increasing in $\lambda$, $\eta$, $\mu$, and $g_N$ provided that $A(\mu, q) < \delta$.*

**Proof.** The characterizing equation (B2) follows from the planner's optimality condition (7). The comparative statics are immediate from (B2). ■

The analogue of Proposition 2 also holds:

**Proposition B.2.** *For all $t > 0$:*

1. *$X(\mu(t), q(t))$ is decreasing in $g_O$.*

2. *There exists an earliest time $\bar{t} < \infty$ such that $X(\mu(t), q(t))$ is decreasing in $g_N$ if $t > \bar{t}$. The time $\bar{t}$ is decreasing in $g_N$.*

3. *Adoption falls to zero as $g_N$ approaches $\rho$, i.e., $\lim_{g_N \uparrow \rho} X(\mu(t), q(t)) = 0$.*

Comparative statics for the evolution of the productivity threshold $A(\mu, q)$ over time are less tractable than for the damage threshold $L(\mu, q)$ in the benchmark model. The following proposition provides some guidance about $\dot{A}(\mu,q)$ and $\ddot{A}(\mu,q)$ for the limiting case in which the new and old technologies grow at the same rate.

**Proposition B.3.** *When $g = g_O = g_N$:*

1. *$\dot{A}(\mu,q)$ is negative and increasing in $g$.*
2. *There exists a posterior $\hat{\mu} \in (0,1/2)$ such that if $\mu \leq \hat{\mu}$, $\ddot{A}(\mu,q)$ is positive.*

**Proof.** When $g = g_O = g_N$, the characterizing equation (B2) becomes

$$A(\mu,q) = \frac{1}{1+\frac{\rho-g}{\mu\lambda\eta}}\delta + \exp(-q).$$

The quality gap $q$ is constant since $g = g_O = g_N$. Differentiating in $t$ then yields

$$\dot{A}(\mu,q) = \dot{\mu}\frac{\frac{\rho-g}{\lambda\eta}}{\left(\mu+\frac{\rho-g}{\lambda\eta}\right)^2}\delta,$$

$$\ddot{A}(\mu,q) = \left[\ddot{\mu} - 2\dot{\mu}^2\frac{1}{\mu+\frac{\rho-g}{\lambda\eta}}\right]\frac{\frac{\rho-g}{\lambda\eta}}{\left(\mu+\frac{\rho-g}{\lambda\eta}\right)^2}\delta.$$

Clearly $\dot{A}(\mu,q) < 0$ because $\dot{\mu} < 0$. Using the equations $\dot{\mu} = -\lambda\mu(1-\mu)$ and $\ddot{\mu} = -\lambda\dot{\mu}(1-2\mu)$, we observe that $\ddot{A}(\mu,q) > 0$ iff

$$1-2\mu > 2\frac{\mu(1-\mu)}{\mu+\frac{\rho-g}{\lambda\eta}}.$$

This inequality is violated at $\mu = 1/2$, but it is satisfied at $\mu = 0$. Hence there exists a cutoff $\hat{\mu} \in (0,1/2)$ such that it is satisfied for $\mu \leq \hat{\mu}$. ■

**Corollary B.1.** *If $g = g_O = g_N$ and $\mu \in (0,\hat{\mu}]$, adoption is concave over time: $\ddot{X}(\mu,q) < 0$.*

These results imply that learning dynamics favor concave adoption over time when sectors are heterogeneous according to comparative advantage, in contrast to the case with heterogeneous damages considered in the main text.

## B.3 Constant Damages

In this section, we assess the role of the assumption that post-disaster damages scale with quality $Q_N$ by revisiting the analysis of Section II under an alternative assumption: Post-disaster damages in sector $i$ are a fixed constant $\Delta_i \geq 0$. In this case, the planner's HJB equations (4, 5) are still valid, but total damages $D(x)$ are now independent of $Q$ and satisfy

$$D(x) = \int_0^1 x_i \Delta_i di.$$

The planner uses technology $N$ in sector $i$ after the disaster iff $\bar{x}_i = 1$ or $\alpha_i Q_N - \Delta_i > Q_O$. If the disaster strikes when the quality vector is $Q$ and the technology choice in sector $i$ is unconstrained, the planner uses technology $O$ for a time period of length $\bar{T}(Q, g, \Delta_i)$, after which she switches to technology $N$. The time period $\bar{T}(Q, g, \Delta_i)$ is equal to zero if $\alpha_i Q_N - \Delta_i \geq Q_O$, and otherwise it is the unique solution to the equation

$$\alpha_i Q_N \exp\left(g_N \bar{T}(Q, g, \delta_i)\right) - \Delta_i = Q_O \exp\left(g_O \bar{T}(Q, g, \delta_i)\right).$$

The solution always exists and is unique since $g_N > g_O$.

By the same argument as in Section II.A, technology $N$ is used in sector $i$ before the disaster if the increase in flow output $\alpha_i Q_N - Q_O$ dominates the expected loss due to the disaster. The latter is the product of the expected arrival rate of the disaster $\mu\lambda$, the probability of irreversibility $\eta_i$, and the difference between the discounted value of net output when technology choice is unconstrained and when it is constrained to technology $N$. If the technology choice in sector $i$ is unconstrained after the disaster, the sector produces discounted net output

$$\int_0^{\bar{T}(Q,g,\delta_i)} \exp(-\rho t) \exp(g_O t) Q_O dt + \int_{\bar{T}(Q,g,\delta_i)}^{\infty} \exp(-\rho t)\left[\alpha_i \exp(g_N t) Q_N - \Delta_i\right] dt.$$

When constrained to technology $N$, the sector's discounted net output is

$$\int_0^{\infty} \exp(-\rho t)\left[\alpha_i \exp(g_N t) Q_N - \Delta_i\right] dt.$$

We then that it is optimal to use technology $N$ in sector $i$ before the disaster iff

$$\text{(B3)} \qquad \alpha_i Q_N - Q_O > \mu\lambda\eta_i \int_0^{\bar{T}(Q,g,\delta_i)} \exp(-\rho t)\left\{\exp(g_O t) Q_O - \left[\alpha_i \exp(g_N t) Q_N - \Delta_i\right]\right\} dt.$$

This optimality condition is analogous to (7) in the benchmark model, but with three differences. First, we have not explicitly integrated the integral in (B3) as we have in (7). Second, in (B3) the fixed damages $\Delta_i$ replace the quality-dependent damages $Q_N \delta_i$ in (7). Finally, with quality-independent damages $\Delta_i$ it is always optimal to use technology $N$ at some point after

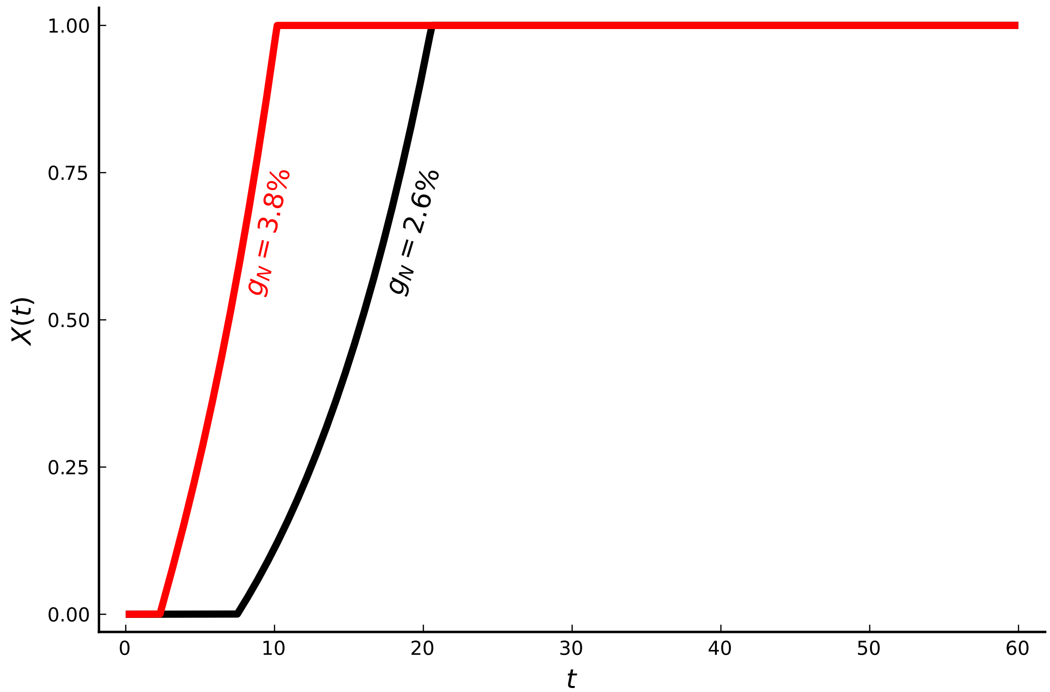


Figure B.1: Adoption curves $X(t) \equiv X(\mu(t), q(t))$ for different values of $g_N$. The parameterization is the same as in Figure 1, but with $\underline{\Delta} = 1$ and $\bar{\Delta} = 5$.

the disaster in sector $i$: $\bar{T}(Q, g, \delta_i) < \infty$. This contrasts with the benchmark model, in which the assumption $\alpha_i \leq \delta_i$ implies that the planner will always use technology $O$ after the disaster when possible.

Suppose as in Section II.B that $\alpha_i$ and $\eta_i$ are constant across sectors. The following proposition is analogous to Proposition 1 in Section II.B. It demonstrates that optimal technology choices can be described using a damage threshold $L(\mu, Q)$ and provides comparative statics.

**Proposition B.4.** *It is socially optimal to use technology $N$ in sector $i$ before the disaster iff $\Delta_i < L(\mu, Q)$, where $L(\mu, Q)$ is the unique solution to the equation*

$$\text{(B4)} \qquad \alpha Q_N - Q_O = \mu\lambda\eta \int_0^{\bar{T}(Q,g,L(\mu,Q))} \exp(-\rho t)\{\exp(g_O t) Q_O - [\alpha \exp(g_N t) Q_N - L(\mu, Q)]\} \, dt.$$

*$L(\mu, Q)$ (and thus $X(\mu, Q)$) is strictly increasing in $\alpha$, $Q_N$, and $g_N$ and strictly decreasing in $g_O$, $\lambda$, $\mu$, and $Q_O$.*

We omit the proof details, because the argument is almost identical to the proof of Proposition C.1 in Appendix C.2. This proposition demonstrates that, when damages from technology $N$ do not scale with its quality $Q_N$, optimal adoption is increasing in the growth rate $g_N$. This contrasts with the corresponding result in Proposition 1, demonstrating that the assumption of proportional damages has significant implications for optimal adoption. We argue that many of the conjectured dangers of (generative) AI more naturally correspond to the case in which damages scale with the capabilities (quality) of rapidly improving models.

We illustrate these results in Figure B.1. We modify the calibration of Figure 1 only by assuming constant damages $\Delta_i$ uniformly distributed over $\left[\underline{\Delta}, \bar{\Delta}\right]$, where $\underline{\Delta} = 1$ and $\bar{\Delta} = 5$, and by initializing $Q(0) = (1, 1)$. As a result, the initial value of the damages in each sector is the same as in the quantitive example in the main text, as is the quality gap $q(0) = 0$. Consistent with Proposition B.4, we observe that adoption is increasing in the growth rate $g_N$. Moreover, adoption is much faster than in Figure 1 because (potential) damages do not increase over time as technology $N$ improves.

# C Analysis with Unrestricted Damages

In this part of the Appendix, we analyze the benchmark model without assuming large damages (6).

## C.1 Socially Optimal Technology Choice

As described in the main text, the planner uses technology $N$ after the disaster iff $\bar{x}_i = 1$ or $(\alpha_i - \gamma_i) Q_N > Q_O$. Letting $q = \log(Q_N / Q_O)$ denote the log quality gap between the technologies, we can equivalently define a threshold gap $q_i$ such that the planner uses technology $N$ after the disaster iff $\bar{x}_i = 1$ or $q \geq q_i$:

$$q_i = \begin{cases} -\log(\alpha_i - \delta_i) & \text{if } \alpha_i > \delta_i, \\ \infty & \text{else.} \end{cases} \tag{C1}$$

At the onset of the disaster, if $q < q_i$ the planner optimally reverts to using technology $O$ in sector $i$ if possible. If $q_i < \infty$, the planner eventually uses technology $N$ again when it attains a sufficiently large lead over technology $O$.

With this characterization, we can directly integrate the post-disaster HJB equation (5) and take expectations with respect to $\bar{x}$:

$$\begin{aligned} \mathbb{E}[W(\bar{x}, Q) | x] = \int_0^1 (1 - x_i \eta_i) \Bigg\{ \Bigg[ 1 - \exp\left( -\frac{\rho - g_O}{g_N - g_O} (q_i - q)_+ \right) \Bigg] \frac{1}{\rho - g_O} Q_O \\ + \exp\left( -\frac{\rho - g_N}{g_N - g_O} (q_i - q)_+ \right) \frac{\alpha_i - \delta_i}{\rho - g_N} Q_N \Bigg\} + x_i \eta_i \frac{\alpha_i - \delta_i}{\rho - g_N} Q_N \, di. \end{aligned}$$

Here we use the notation $(q_i - q)_+ = \max\{q_i - q, 0\}$. Considering the planner's problem before the disaster (4), we observe that it is optimal to use technology $N$ in sector $i$ iff

$$\begin{aligned} \alpha_i Q_N - Q_O > \mu \lambda \eta_i \Bigg\{ \Bigg[ 1 - \exp\left( -\frac{\rho - g_O}{g_N - g_O} (q_i - q)_+ \right) \Bigg] \frac{1}{\rho - g_O} Q_O \\ - \Bigg[ 1 - \exp\left( -\frac{\rho - g_N}{g_N - g_O} (q_i - q)_+ \right) \Bigg] \frac{\alpha_i - \delta_i}{\rho - g_N} Q_N \Bigg\}. \end{aligned} \tag{C2}$$

This optimality condition differs from (7) because the discounted future net output from using technology $O$ at the time of the disaster now accounts for the possibility that technology $N$ is used after the quality gap $q$ exceeds $q_i$.

## C.2 Comparative Statics for Socially Optimal Adoption

Suppose as in Section II.B that $\alpha_i$ and $\eta_i$ are constant across sectors, but make no assumption about the ranking between $\delta_i$ and $\alpha$. Let $\bar{q}(\delta_i) = q_i$ denote the quality gap above which it is optimal to use technology $N$ in sector $i$ after the disaster (C1), making explicit the dependence on $\delta_i$. The following proposition shows that optimal technology choices can be described using a damage threshold $L(\mu, q)$ and provides comparative statics, generalizing Proposition 1 from Section II.B.

**Proposition C.1.** *It is socially optimal to use technology $N$ in sector $i$ before the disaster iff $\delta_i < L(\mu, q)$, where $L(\mu, q)$ is the unique solution to the equation*

$$\text{(C3)} \quad \alpha - \exp(-q) = \mu\lambda\eta \left\{ \left[ 1 - \exp\left( -\frac{\rho - g_O}{g_N - g_O} \left( \bar{q}(L(\mu, q)) - q \right)_+ \right) \right] \frac{1}{\rho - g_O} \exp(-q) - \left[ 1 - \exp\left( -\frac{\rho - g_N}{g_N - g_O} \left( \bar{q}(L(\mu, q)) - q \right)_+ \right) \right] \frac{\alpha - L(\mu, q)}{\rho - g_N} \right\}.$$

*$L(\mu, q)$ (and thus $X(\mu, q)$) is strictly increasing in $\alpha$ and $q$ and strictly decreasing in $g_O$, $\lambda$, and $\mu$. It is strictly decreasing in $g_N$ if $L(\mu, q) > \alpha$ and strictly increasing in $g_N$ if $L(\mu, q) < \alpha$.*

**Proof.** Throughout the proof, we suppress the arguments of the damage threshold $L(\mu, q)$ to simplify notation. The results described in the proposition are easier to prove if we re-write the discounted values on the right-hand side of (C3) as integrals over time. To do this, given a quality gap $q$, let $\bar{T}(q, g, \delta)$ denote the length of time after the disaster during which it is optimal to use technology $O$ instead of technology $N$ in a sector with damages $\delta$:

$$\bar{T}(q, g, \delta) = \begin{cases} \max\left\{ \frac{-\log(\alpha - \delta) - q}{g_N - g_O}, 0 \right\} & \text{if } \alpha > \delta, \\ \infty & \text{else.} \end{cases}$$

If the sector is not constrained to technology $N$, its discounted net output after the disaster is

$$\text{(C4)} \quad \int_0^{\bar{T}(q,g,\delta)} \exp(-\rho t) \exp(g_O t) Q_O \, dt + \int_{\bar{T}(q,g,\delta)}^{\infty} \exp(-\rho t) \exp(g_N t)(\alpha - \delta) Q_N \, dt.$$

Similarly, its discounted net output when constrained to technology $N$ is

$$\text{(C5)} \quad \int_0^{\infty} \exp(-\rho t) \exp(g_N t)(\alpha - \delta) Q_N \, dt.$$

The bracketed term in (C3) is the difference between the previous two terms above for the marginal sector (with $\delta = L$), divided by $Q_N$. The right-hand side of (C3) can then be written

$$\text{RHS} = \mu\lambda\eta \int_0^{\bar{T}(q,g,L)} \exp(-\rho t)\left[\exp(g_O t)\exp(-q) - \exp(g_N t)(\alpha - L)\right] dt.$$

We first demonstrate that, when $\alpha > \exp(-q)$ so that technology $N$ is more productive than technology $O$, there always exists a unique solution $L$ to (C3). We observe that RHS is continuous in $L$, equals zero when $L \leq \alpha - \exp(-q)$, and limits to infinity as $L \to \infty$. Moreover, RHS is strictly increasing in $L$ when $L > \alpha - \exp(-q)$: This condition implies $\bar{T}(q, g, L) > 0$, and we can differentiate RHS to find

$$\begin{aligned}
\frac{\partial \text{RHS}}{\partial L} &= \mu\lambda\eta \exp\left(-\rho \bar{T}\right)\left[\exp\left(g_O \bar{T}\right)\exp(-q) - \exp\left(g_N \bar{T}\right)(\alpha - L)\right] \frac{\partial \bar{T}}{\partial L} \\
&\quad + \mu\lambda\eta \int_0^{\bar{T}(q,g,L)} \exp(-\rho t)\exp(g_N t)\, dt \\
&= \mu\lambda\eta \int_0^{\bar{T}(q,g,L)} \exp(-\rho t)\exp(g_N t)\, dt \\
&> 0.
\end{aligned}$$

Note that the second equality holds by the Envelope Theorem: $\bar{T}$ maximizes the discounted net output from the marginal sector after the disaster, assuming its technology choice is unconstrained. As a result RHS does not vary locally with respect to $\bar{T}$ ($\partial \text{RHS}/\partial \bar{T} = 0$). Given these properties of RHS, the Intermediate Value Theorem guarantees a unique solution $L$ to (C3) when $\alpha > \exp(-q)$. Moreover, it follows from the optimality condition (C2) that it is socially optimal to use technology $N$ in sector $i$ before the disaster iff $\delta_i < L(\mu, q)$.

The comparative statics for the damage threshold $L$ follow from the Implicit Function Theorem. Holding $L$ fixed, we immediately observe that RHS is decreasing in $\alpha$ and increasing in $\mu$, $\lambda$, and $\eta$. Differentiating with respect to $q$, $g_O$, and $g_N$ yields

$$\begin{aligned}
\frac{\partial \text{RHS}}{\partial q} &= -\mu\lambda\eta \int_0^{\bar{T}} \exp(-\rho t)\exp(g_O t)\exp(-q)\, dt, \\
\frac{\partial \text{RHS}}{\partial g_O} &= \mu\lambda\eta \int_0^{\bar{T}} \exp(-\rho t)\exp(g_O t)\exp(-q)\, t\, dt, \\
\frac{\partial \text{RHS}}{\partial g_N} &= -\mu\lambda\eta \int_0^{\bar{T}} \exp(-\rho t)\exp(g_N t)(\alpha - L(\mu, q))\, dt.
\end{aligned}$$

These expressions imply that RHS is decreasing in $q$, increasing in $g_O$, and decreasing (increasing) in $g_N$ iff $\alpha > (<) L(\mu, q)$. Collecting these results, the Implicit Function Theorem delivers the comparative statics stated in the proposition. ■

The proposition demonstrates that almost all comparative statics from Proposition 1 hold without the assumption that social damages always exceed output from technology $N$ after the disaster ($\alpha_i \leq \delta_i$). However, the comparative static with respect to $g_N$ is sensitive to this assumption. When damages in the marginal sector exceed output ($L(\mu, q) > \alpha$), the damage threshold is decreasing in $g_N$ as in Proposition 1. When damages in the marginal sector are below output ($L(\mu, q) < \alpha$), the damage threshold is instead increasing in $g_N$.

The following proposition generalizes Proposition 2 to provide full comparative statics for adoption with respect to the growth rates $g_O$ and $g_N$, including both the direct effects described in Proposition C.1 and the indirect effects through the state $(\mu(t), q(t))$.

**Proposition C.2.** *Suppose $\alpha_i$ and $\eta_i$ are constant across sectors. For all $t$ with $L(\mu(t), q(t)) < \alpha$:*

1. *$X(\mu(t), q(t))$ is decreasing in $g_O$.*
2. *$X(\mu(t), q(t))$ is increasing in $g_N$.*
3. *If $q(0)$ is sufficiently low and $X(\mu(t), q(t)) < F(\alpha)$, $X(\mu(t), q(t))$ is bounded strictly below $F(\alpha)$ as $g_N$ approaches $\rho$, i.e., $\lim_{g_N \uparrow \rho} X(\mu(t), q(t)) < F(\alpha)$.*

**Proof.** The first two results follow from Proposition C.1 after noting that the damage threshold $L$ is increasing in the quality gap $q$, and in turn the quality gap $q(t)$ at time $t$ is decreasing in $g_O$ and increasing in $g_N$. The final result of the proposition follows by taking the limit $g_N \uparrow \rho$ in (C3). More precisely, recall from the optimality condition (C2) that it is socially optimal to use technology $N$ in sector $i$ at time $t$ before the disaster only if

$$
\text{(C6)} \qquad \alpha - \exp(-q(t)) \geq \mu(t)\lambda\eta \left\{ \left[ 1 - \exp\left( -\frac{\rho - g_O}{g_N - g_O} (q_i - q(t))_+ \right) \right] \frac{1}{\rho - g_O} \exp(-q(t)) - \left[ 1 - \exp\left( -\frac{\rho - g_N}{g_N - g_O} (q_i - q(t))_+ \right) \right] \frac{\alpha - \delta_i}{\rho - g_N} \right\}.
$$

Just as in the proof of Proposition C.1, the right-hand side of this inequality can be written

$$
\text{RHS} = \mu(t)\lambda\eta \int_0^{\bar{T}(q(t), g, \delta_i)} \exp(-\rho s) \left[ \exp(g_O s) \exp(-q(t)) - \exp(g_N s)(\alpha - \delta_i) \right] ds,
$$

For $g_N < \rho$, this function is continuous and nondecreasing in $\delta_i$. At $\delta_i = \alpha$, it takes the value

$$
\text{RHS} = \frac{\mu(t)\lambda\eta}{\rho - g_O} \exp(-q(t)).
$$

In the limit as $g_N \uparrow \rho$, (C6) implies that it is socially optimal to use technology $N$ in sector $i$ with damages $\delta_i = \alpha$ only if

$$\alpha \exp(q(0) + (\rho - g_O)t) - 1 \geq \frac{\mu(t)\lambda\eta}{\rho - g_O}.$$

For $t$ fixed, this inequality is violated for $q(0)$ sufficiently small (but potentially still positive). Since RHS is continuous in $\delta_i \leq \alpha$, this implies that $L(\mu(t), q(t))$ must remain bounded strictly below $\alpha$. If $X(\mu(t), q(t)) < F(\alpha)$, this immediately implies that $X(\mu(t), q(t))$ is bounded strictly below $F(\alpha)$. ■

Notably, in the limit $g_N \uparrow \rho$ adoption does not tend to either of the extreme values 0 or $F(\alpha)$, in contrast to the corresponding result in Proposition 2. This holds because, for any sector $i$ with $\delta_i < \alpha$, the discounted net output after the disaster tends to infinity as $g_N \uparrow \rho$ regardless of whether the sector is constrained to use technology $N$ after the disaster. However, the *difference* between the discounted net output when unconstrained and the discounted net output when constrained tends to a finite limit. Socially optimal technology choices before the disaster depend on this difference (see C2), so provided that $\delta_i$ is sufficiently close to $\alpha$ and the initial quality gap $q(0)$ sufficiently low, it can remain optimal to delay using technology $N$ in sector $i$ before the disaster even when $g_N \uparrow \rho$.

We illustrate these results in Figure C.1 by depicting adoption curves for a stylized parameterization of the model. We modify the calibration of Figure 1 only by assuming that the distribution of damages $\delta_i$ is uniform over $[0,5]$ instead of $[1,5]$. Technology choices for sectors with $\delta_i \in [1,5]$ are exactly as in Section II, and since these sectors comprise $5/6$ of all sectors in this calibration, the adoption curves in Figure C.1 when $X(t) \geq 5/6$ are identical to the adoption curves in Figure 1.

When instead $X(t) \in (0, 1/6)$, the sectors adopting technology $N$ produce positive net output after the disaster, so the analysis in this appendix becomes relevant. In this region, we observe that adoption is increasing in $g_N$, consistent with Proposition C.2.

## C.3 Equilibrium Technology Choice

Using the same derivations as for the optimal technology choice, firm $i$ uses technology $N$ after the disaster iff $\bar{x}_i = 1$ or $q \geq \tilde{q}_i$, where

$$\tilde{q}_i = \begin{cases} -\log(\alpha_i - \gamma_i) & \text{if } \alpha_i > \gamma_i, \\ \infty & \text{else.} \end{cases}$$

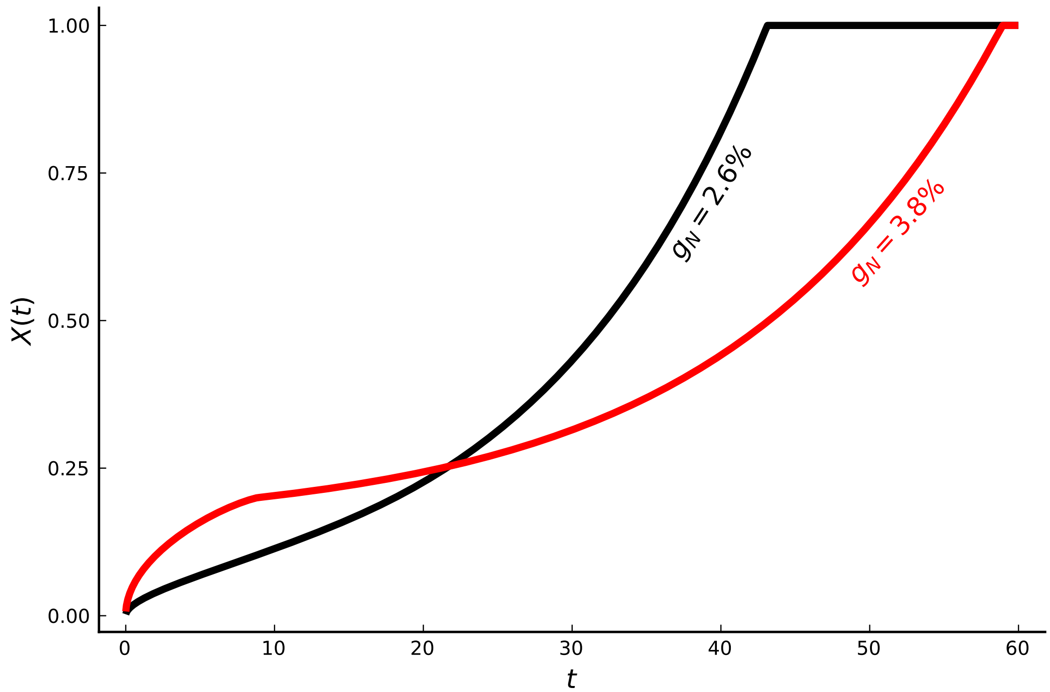


Figure C.1: Adoption curves $X(t)$ for different values of $g_N$. The parameterization is the same as in Figure 1, but with $\underline{\delta} = 0$.

Note that $\tilde{q}_i \leq q_i$ since $\gamma_i \leq \delta_i$. This implies that the private firm returns to using technology $N$ more quickly after the disaster than the planner. Integrating the firm's post-disaster HJB equation (10) and taking expectations with respect to $\bar{x}_i$ yields

$$\mathbb{E}\left[\Phi_i(\bar{x}_i, Q) | x_i\right] = (1 - x_i \eta_i) \left\{ \left[1 - \exp\left(-\frac{\rho - g_O}{g_N - g_O}(\tilde{q}_i - q)_+\right)\right] \frac{1}{\rho - g_O} Q_O + \exp\left(-\frac{\rho - g_N}{g_N - g_O}(\tilde{q}_i - q)_+\right) \frac{\alpha_i - \gamma_i}{\rho - g_N} Q_N \right\} + x_i \eta_i \frac{\alpha_i - \gamma_i}{\rho - g_N} Q_N.$$

It is then privately optimal to use technology $N$ in sector $i$ before the disaster iff

$$\text{(C7)} \qquad \alpha_i Q_N - Q_O > \mu \lambda \eta_i \left\{ \left[1 - \exp\left(-\frac{\rho - g_O}{g_N - g_O}(\tilde{q}_i - q)_+\right)\right] \frac{1}{\rho - g_O} Q_O - \left[1 - \exp\left(-\frac{\rho - g_N}{g_N - g_O}(\tilde{q}_i - q)_+\right)\right] \frac{\alpha_i - \gamma_i}{\rho - g_N} Q_N \right\}.$$

We observe two differences between this condition and the planner's optimality condition (C2), First, as in the main text, private damages $\gamma_i$ appear in (C7) instead of the social damages that appear in (C2). Second, the firm begins using technology $N$ more quickly after the disaster than the planner ($\tilde{q}_i \leq q_i$). Both effects tend to reduce the net private cost of irreversibility and incentivize the firm to use technology $N$ more often than the planner before the disaster.

**Lemma C.1.** *If the social planner uses technology $N$ in sector $i$ in state $(\mu, Q)$ before the disaster, then so does firm $i$.*

**Proof.** The statement holds provided that firm $i$'s opportunity cost to using technology $N$ instead of technology $O$ at the time of the disaster is smaller than the planner's opportunity cost:

$$\left[1-\exp\left(-\frac{\rho-g_O}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{1}{\rho-g_O}Q_O-\left[1-\exp\left(-\frac{\rho-g_N}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{\alpha_i-\gamma_i}{\rho-g_N}Q_N$$
$$\leq\left[1-\exp\left(-\frac{\rho-g_O}{g_N-g_O}(q_i-q)_+\right)\right]\frac{1}{\rho-g_O}Q_O-\left[1-\exp\left(-\frac{\rho-g_N}{g_N-g_O}(q_i-q)_+\right)\right]\frac{\alpha_i-\delta_i}{\rho-g_N}Q_N.$$

Replacing $\gamma_i$ with $\delta_i$ yields the intermediate inequality

$$\left[1-\exp\left(-\frac{\rho-g_O}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{1}{\rho-g_O}Q_O-\left[1-\exp\left(-\frac{\rho-g_N}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{\alpha_i-\gamma_i}{\rho-g_N}Q_N$$
$$\leq\left[1-\exp\left(-\frac{\rho-g_O}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{1}{\rho-g_O}Q_O-\left[1-\exp\left(-\frac{\rho-g_N}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{\alpha_i-\delta_i}{\rho-g_N}Q_N.$$

Optimality of $q_i$ in the planner's problem after the disaster yields the remaining inequality

$$\left[1-\exp\left(-\frac{\rho-g_O}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{1}{\rho-g_O}Q_O-\left[1-\exp\left(-\frac{\rho-g_N}{g_N-g_O}(\tilde{q}_i-q)_+\right)\right]\frac{\alpha_i-\delta_i}{\rho-g_N}Q_N$$
$$\leq\left[1-\exp\left(-\frac{\rho-g_O}{g_N-g_O}(q_i-q)_+\right)\right]\frac{1}{\rho-g_O}Q_O-\left[1-\exp\left(-\frac{\rho-g_N}{g_N-g_O}(q_i-q)_+\right)\right]\frac{\alpha_i-\delta_i}{\rho-g_N}Q_N.$$

■

# D Other Issues in Regulation

In this part of the Appendix, we explore second-best tax regulation schemes when private and social damages are not positively affiliated, and we provide additional details about optimal regulatory sandboxes and discuss their advantages relative to sector-independent taxes.

## D.1 Second-Best Tax Regulation

Aside from the special case in which social and private damages are positively affiliated, a sector-independent tax cannot implement the optimal technology choices in equilibrium. More generally, use taxes can allow the planner to improve upon laissez-faire technology choices even when optimal ones cannot be implemented. Suppose as in Proposition 8 that $\alpha_i$ and $\eta_i$ are constant across sectors, but make no assumptions on the joint distribution of $\delta_i$ and $\gamma_i$. In each state $(\mu, Q)$ before the disaster, the planner chooses the use tax $\tau(\mu, Q)$ to maximize output less the expected discounted social cost from the disaster:

$$\max_{\tau} \int_0^1 \left\{ \left(1 - x\left(\mu, Q, \gamma_i, \tau\right)\right) \left[ Q_O + \mu\lambda\eta \frac{1}{\rho - g_O} Q_O \right] + x\left(\mu, Q, \gamma_i, \tau\right) \left[ \alpha Q_N + \mu\lambda\eta \frac{\alpha - \delta_i}{\rho - g_N} Q_N \right] \right\} di.$$

Here $x\left(\mu, Q, \gamma_i, \tau\right)$ describes the equilibrium technology choice for firm $i$ when subject to the tax:

$$x\left(\mu, Q, \gamma_i, \tau\right) = \begin{cases} 1 & \text{if } \alpha Q_N - Q_O - \tau > \mu\lambda\eta \left[ \frac{1}{\rho - g_O} Q_O - \frac{\alpha - \gamma_i}{\rho - g_N} Q_N \right], \\ 0 & \text{else.} \end{cases}$$

Firms adopt technology $N$ in order of increasing $\gamma_i$, so we can equivalently assume that the planner selects a private damage threshold $\hat{L}(\mu, q)$ such that firm $i$ uses technology $N$ iff $\gamma_i < \hat{L}(\mu, q)$. The optimal threshold trades off flow consumption against the expected social cost of the disaster. When interior, it satisfies

$$\alpha Q_N - Q_O = \mu\lambda\eta \left[ \frac{1}{\rho - g_O} Q_O - \frac{\alpha - \bar{\delta}\left(\hat{L}(\mu, q)\right)}{\rho - g_N} Q_N \right]. \tag{D1}$$

Here $\bar{\delta}(\gamma) = \mathbb{E}\left[\, \delta_i \middle| \gamma_i = \gamma \right]$ is the average social damages across all firms with private damages $\gamma$. The optimality condition (D1) is analogous to the original optimality condition (7), but it replaces a single sector's social damages $\delta_i$ with the expectation $\bar{\delta}(\gamma)$. The planner's problem is concave iff $\bar{\delta}(\gamma)$ is increasing, in which case an interior solution can be optimal. If, for example,

$\bar{\delta}(\gamma)$ is decreasing, then the planner cannot incentivize sectors with low social damages to use technology $N$ while sectors with high social damages use technology $O$. As a result, the planner chooses $\hat{L}(\mu,q)=0$ (no use of $N$) or $\hat{L}(\mu,q)=\infty$ (full use of $N$). The latter is optimal when

$$\alpha Q_N - Q_O > \mu\lambda\eta\left[\frac{1}{\rho-g_O}Q_O - \frac{\alpha-\mathbb{E}[\delta_i]}{\rho-g_N}Q_N\right].$$

## D.2 Analysis of Sandbox Regulation

Proposition 9 in the main text demonstrates that it is generally optimal for the planner to implement a regulatory sandbox with a strictly positive wait time $\hat{T}$. The optimal wait time $\hat{T}$ must satisfy the following interior first-order condition, which is derived in the proof of the proposition in Appendix A:

$$\text{(D2)} \qquad 0 = -\int_{\delta_i \geq \hat{\delta}} x\left(\mu,q,\gamma_i\right)\left\{\alpha Q_N - Q_O - \mu\lambda\eta\left[\frac{1}{\rho-g_O}Q_O - \frac{\alpha-\delta_i}{\rho-g_N}Q_N\right]\right\} di.$$

Here the state $(\mu,Q)$ is evaluated at the optimal time $\hat{T}$, and $x\left(\mu,q,\gamma_i\right)=1$ iff sector $i$ would use technology $N$ in the laissez-faire equilibrium. Two forces determine the optimal wait time $\hat{T}$: If sector $i$ is above the threshold ($\delta_i \geq \hat{\delta}$) and would inefficiently use technology $i$ at time $\hat{T}$, its laissez-faire technology choice would decrease social welfare, favoring a longer wait time:

$$x\left(\mu,q,\gamma_i\right)\left\{\alpha Q_N - Q_O - \mu\lambda\eta\left[\frac{1}{\rho-g_O}Q_O - \frac{\alpha-\delta_i}{\rho-g_N}Q_N\right]\right\} < 0.$$

If sector $i$ would instead efficiently use technology $i$ at time $\hat{T}$, its laissez-faire technology choice would increase social welfare, favoring a shorter wait time.[14]

We can similarly derive the following interior first-order condition for the optimal threshold $\hat{\delta}$, keeping $\hat{T}$ fixed:

$$0 = \int_0^{\hat{T}} \exp(-\rho t)\int_{\delta_i=\hat{\delta}} x\left(\mu,q,\gamma_i\right)\left\{\alpha Q_N - Q_O - \mu\lambda\eta\left[\frac{1}{\rho-g_O}Q_O - \frac{\alpha-\hat{\delta}}{\rho-g_N}Q_N\right]\right\} dt.$$

If the threshold $\hat{\delta}$ is too high, a large fraction of sectors $i$ face no restrictions on their technology choices, and they subtract too much from social welfare between $t=0$ and $t=\hat{T}$ as they begin using the new technology too quickly. If $\hat{\delta}$ is too low, then too many sectors $i$ are forced to use technology $O$ between $t=0$ and $t=\hat{T}$, foregoing the benefits of using technology $N$ in these

[14]As this intuition suggests, it is straightforward to verify that, under the assumptions of Proposition 9, the optimal wait time $\hat{T}$ is nondecreasing in $\hat{\delta}$.

sectors when it is efficient to do so. This analysis demonstrates that the optimal parameters $(\hat{\delta}, \hat{T})$ are chosen to resolve a trade-off between restricting early use of the new technology in sectors where expected damages are large, while allowing broad use later as the probability of a disaster falls and the quality gap grows.

We conclude this section by observing that regulatory sandboxes are likely to dominate (or complement) sector-independent taxes when the order of adoption differs substantially between the equilibrium and social optimum. For example, suppose that private and social damages are negatively affiliated: $\gamma_i = \kappa(\delta_i)$, where $\kappa$ is strictly decreasing. Then Proposition 8 implies that, for any sector-independent tax $\tau(\mu, Q)$, the order in which sectors adopt the new technology in equilibrium is exactly the opposite of the optimal order. Moreover, the analysis in Appendix D.1 implies that the optimal sector-independent tax is such that there exists a time $\hat{T}$ before which no sector uses technology $N$ and after which every sector uses technology $N$. This time is characterized by the equation

$$\alpha Q_N(\hat{T}) - Q_O(\hat{T}) = \mu(\hat{T})\lambda\eta\left[\frac{1}{\rho - g_O}Q_O(\hat{T}) - \frac{\alpha - \mathbb{E}[\delta_i]}{\rho - g_N}Q_N(\hat{T})\right].$$

These technology choices can also be implemented using the sandbox policy with threshold $\hat{\delta} = \underline{\delta}$ and wait time $\hat{T}$. Hence a regulatory sandbox can achieve weakly greater social welfare than any sector-independent tax when the misalignment in the order of adoption between the equilibrium and the social optimum is severe.