# Panics and Early Warnings

Deepal Basak

*Indiana University*

Zhen Zhou

*Tsinghua University*

We study optimal adversarial information design in a dynamic regime change game. Agents decide when to attack, if at all. We assume (1) delay incurs a continuous cost and (2) agents doubt the correctness of their actions. The game may end in a "disaster" due to weak fundamentals or panic—agents attacking despite sound fundamentals. We propose a "timely disaster alert" that promptly warns about impending disasters, making waiting for and following the alert the unique rationalizable strategy, thereby eliminating panic. We relate this optimal policy to early-warning systems such as bank stress tests and debt sustainability analysis.

## I. Introduction

A disaster may happen even when the fundamentals do not seem to warrant one. For example, if creditors anticipate that a bank will face a negative shock in the future, such as borrowers in a certain industry defaulting

This paper was previously circulated under the title "Timely Persuasion." We thank Emiliano Catonini, Christophe Chamley, Joyee Deb, Douglas Gale, Rick Harbaugh, Ju Hu, Nicolas Inostroza, Aaron Kolb, Jiangtao Li, Stephen Morris, Mariann Ollár, Alessandro Pavan, David Pearce, Antonio Penta, Debraj Ray, Edouard Schaal, Ennio Stacchetti, Colin Stewart, Martin Szydlowski, Laura Veldkamp, Xingye Wu, and Ming Yang, as well as the seminar participants at the Econometric Society meetings, Stony Brook University, Peking University, Tsinghua University, Fudan University, New York University Shanghai, NUS, UPF, SMU, University of Washington, University of Toronto, University of Kansas, Indiana University's Kelley

Electronically published June 5, 2025

*Journal of Political Economy*, volume 133, number 7, July 2025.

https://doi.org/10.1086/734875

on their loans, they may start withdrawing their money from the bank. This action can weaken the bank and cause a bank run even though the shock may not be severe when it arrives. We call such events *panic*. Panics are not limited to bank runs but can occur in many coordination problems, such as sovereign debt defaults and capital outflow. This paper studies how to design an information disclosure policy that averts panic.

### A. *Regime Change Game*

We consider a canonical regime change game à la Morris and Shin (2003). A mass of agents receive private information (which could be arbitrarily correlated) and decide whether to attack a regime, and the game ends in a disaster if the regime is not strong enough to withstand the attack. The agents prefer to attack if the game ends in a disaster and prefer not to attack if it does not. We introduce a natural timing dimension to the problem: the agents do not move simultaneously but rather endogenously choose when to attack, if at all. We make two assumptions:

(1) Cost: A delayed attack is costly, and the cost is continuously increasing in the duration of waiting.
(2) Doubt: Regardless of their private signals and regardless of what others do, agents are uncertain whether their actions are correct.

A principal aims to prevent a disaster and commits to a dynamic information disclosure policy. The principal anticipates that, for any policy she chooses, the worst outcome will happen. We study the optimal dynamic adversarial information design in this setup.

We construct an optimal disclosure policy called the *timely disaster alert*. The principal promises to send a binary public signal at a future date: trigger the alert or not. She triggers the alert if, and only if, the disaster is inevitable. Note that a disaster may become inevitable because the fundamental is bad or because some agents attacked rather than waited for the disclosure. We say that the alert is timely if the duration of waiting for this alert is sufficiently short. We show that under this policy, in the unique rationalizable strategy, the agents wait for the alert and then follow it, regardless of their private signals. Accordingly, there is no attack unless the fundamental warrants a disaster. In this sense, this policy eliminates panic, and the principal achieves her most preferred outcome.

School of Business, University of Rochester's Simon Business School, University of Illinois and Urbana-Champaign, University of Utah's Eccles School of Business, Ashoka University, Michigan State University, University of Minnesota, and SET (online) for their helpful comments and suggestions. We thank Samyak Jain and Dandan Zhao for their excellent research assistance and the National Science Foundation of China (project no. 72322005) and Tsinghua University Initiative Scientific Research Program for financial support.

Our proposed timely disaster alert resembles an early warning system (EWS) used in practice. For instance, the International Monetary Fund (IMF) and the World Bank regularly conduct forward-looking debt sustainability analysis (DSA), aiming to provide the market with an early warning of sovereign debt distress. After the Great Recession, bank supervisors adopted forward-looking bank stress tests to assess banks' performance under some predicted future shock, and they publicly disclosed whether or not a bank would be able to sustain that adverse shock. Our analysis provides a rationale for adopting an EWS in a fairly general setup, provided that the design of an EWS incorporates the historical market reactions or endogenous attacks and discloses information in a timely manner.[1]

The result may seem surprising, particularly given the well-known result in the literature (which we will discuss shortly) that a principal cannot design an information disclosure policy to eliminate panic in a static regime change game under adversarial selection. In contrast, when there is a natural timing dimension to the problem, the principal can indeed eliminate panic. To elucidate the intuition behind our main result and the source of this contrasting outcome, we first present a simple example involving two players over two periods.[2]

### *B. A Two-Player, Two-Period Example*

There is an unknown state that is bad ($B$), medium ($M$), or good ($G$). Two players share a common belief: $\pi_\theta \in (0, 1)$ is the probability that the state is $\theta \in \{B, M, G\}$. We are interested in the environment where the state is likely medium, but with a small positive probability, it can be good or bad. This assumption ensures that both good and bad equilibria exist, and the information design problem under adversarial selection is nontrivial.

In the first period, each player simultaneously decides whether to take an irreversible action named "attack" ($A$) or "wait" ($W$). If he attacks, he has no more decisions to make, and if he waits, then in period 2, he again chooses whether to attack ($A$) or not ($N$). The players do not see the action of the other player. Each player's payoff is as follows:

(1) If he attacks right away ($A$), he gets 1.
(2) If he waits and then attacks ($WA$), he still gets 1 either when the state is good or when the state is medium and the other player

[1] The forward-looking bank stress test serves as an early warning system to prevent bank panic in practice. In the United States, the Supervisory Capital Assessment Program (SCAP) was conducted in 2009 to restore confidence in the banking sector. The procedures for SCAP were announced on February 10, 2009, and the results were disclosed on May 7, 2009. Several studies (e.g., Peristiani, Morgan, and Savino 2010; Bernanke 2013; Gorton 2015) report that this stress test was effective. For further details, see sec. II.1 of the online appendix.

[2] We thank an anonymous referee for suggesting this example.

does not attack in either period 1 or period 2; otherwise, he gets $1 - c$, where $c \in (0, 1)$ stands for the delay cost of attacking.

(3) If he never attacks (*WN*), he gets 2 if either the state is good or the state is medium and the other player does not attack in either period 1 or period 2 (*A* or *WA*); otherwise, he gets 0.

Table 1 describes how the payoff depends on the state $\theta$ and both players' strategies.[3]

In this game, without any information revealed before period 2, either about the state or about the period 1 play, no one would wait and then attack since *WA* is strictly dominated by *A*. This is because, compared with attacking at $t = 1$, waiting and attacking later does not bring about any benefit, whereas it is strictly worse when $\theta = B$, which is believed to occur with a strictly positive probability $\pi_B > 0$.

Therefore, this dynamic game is reduced to the static game in which the players decide between attacking (*A*) or not attacking (*WN*). If $\pi_B \leq 1/2$, there is a "good" equilibrium where the players do not attack; that is, they play *WN*. However, if $\pi_G \leq 1/2$, there is also a "bad" equilibrium where both players attack; that is, they play *A*. Note that if $\theta \neq B$, then coordinating on not attacking is the efficient outcome. However, under the bad equilibrium, this efficient coordination is not achieved. We refer to this event as *panic*.

We are interested in the adversarial information design problem where a principal adopts a disclosure policy to avert such panic. We say that panic is eliminated if, in all possible equilibria, the players coordinate on not attacking whenever it is efficient to do so (i.e., $\theta \neq B$).

Before introducing our proposed optimal disclosure policy, let us briefly discuss why some straightforward disclosure policies do not work. First, consider the full disclosure policy that reveals the true state before period 1 play. Under this policy, when $\theta = M$, the players face a standard coordination problem. In this case, efficient coordination is possible, but both players attacking after observing $\theta = M$ is clearly an equilibrium. Moreover, instead of fully revealing the state, the principal may only disclose whether the state is "bad" ($\theta = B$) or not ($\theta \neq B$) before period 1. However, as long as $\pi_M \geq \pi_G$, there remains an equilibrium where both players attack after learning $\theta \neq B$. Therefore, these static disclosure policies do not avert panic. In addition to the disclosure before period 1, the principal may want to make a separate announcement, prior to period 2

[3] One interpretation of the payoff is as follows. There is a regime with fundamental state $\theta$. The regime changes if and only if $\theta = B$, or $\theta = M$ and at least one player chooses to attack in either period 1 or period 2. The payoff from attacking is fixed at 1, but a delayed attack incurs a cost $c$ only when the regime changes; and the payoff from not attacking is 0 (2) if the regime changes (does not change).

TABLE 1
PAYOFF MATRIX

| | *A* | *WA* | *WN* |
|---|---|---|---|
| $\theta = B$ | | | |
| *A* | (1, 1) | (1, 1 − *c*) | (1, 0) |
| *WA* | (1 − *c*, 1) | (1 − *c*, 1 − *c*) | (1 − *c*, 0) |
| *WN* | (0, 1) | (0, 1 − *c*) | (0, 0) |
| $\theta = M$ | | | |
| *A* | (1, 1) | (1, 1 − *c*) | (1, 0) |
| *WA* | (1 − *c*, 1) | (1 − *c*, 1 − *c*) | (1, 0) |
| *WN* | (0, 1) | (0, 1) | (2, 2) |
| $\theta = G$ | | | |
| *A* | (1, 1) | (1, 1) | (1, 2) |
| *WA* | (1, 1) | (1, 1) | (1, 2) |
| *WN* | (2, 1) | (2, 1) | (2, 2) |

play, about whether anyone has attacked in period 1. However, this additional information cannot change the fact that both players attacking in period 1 after $\theta = M$ (or $\theta \neq B$) remains an equilibrium. This is because if a player believes that the opponent will take this strategy, then he expects to learn nothing useful by waiting. Therefore, if both players attacking is an equilibrium without this new information, it remains so even with this new information.[4]

We propose a dynamic disclosure policy that combines information about both the fundamental and endogenous attacks in one disclosure. We call this a *disaster alert*. The principal sets a binary disclosure before the players move in period 2, which sends a public message of "alert" ($d = 1$) if (1) $\theta = B$ or (2) $\theta = M$ and the opponent has already attacked, and "no alert" ($d = 0$) otherwise. Note that the (conditionally) dominant strategy is for a player to attack at period 2 if he sees the alert. Under a disaster alert policy, the five possible strategies are (1) *A*, (2) *WAA*, (3) *WAN*, (4) *WNA*, and (5) *WNN*. If a player chooses to wait at $t = 1$ (denoted as *W*), then the first letter following *W* represents his choice following no alert ($d = 0$), and the second one presents her choice following $d = 1$. For instance, *WNA* means to wait at $t = 1$, not attack at $t = 2$ following $d = 0$, and attack at $t = 2$ following $d = 1$. We call the strategy *WNA wait and follow*.

[4] To see this formally, recall that *A* gives a payoff of 1. If the player does not attack right away (after learning $\theta = M$ under full disclosure or $\theta \neq B$ under partial disclosure), then he can play *AA*, *NA*, *AN*, or *NN* (where the first letter signifies the action after seeing that the opponent does not attack and the second letter signifies the action after seeing that the opponent does attack). Suppose he believes that the opponent will play *A*. Then, his payoff from *AA* or *NA* is what he gets from attacking in period 2, which is always lower than 1. On the other hand, his payoff from *AN* or *NN* is what he gets from never attacking. Since (*A*, *A*) is an equilibrium without the period 2 information, the payoff from this strategy must be lower than 1. Therefore, if he believes the opponent will play *A*, he cannot do any better than playing *A* himself, which implies that (*A*, *A*) remains an equilibrium.

Below, we show that if the delay cost is sufficiently small (i.e., $c < \pi_G/(1 - \pi_G)$), then the only strategy that survives iterated elimination of strictly dominated strategies is wait and follow (*WNA*).

In the first round, strategies *WAA*, *WAN*, and *WNN* are eliminated. First, note that attacking at $t = 2$ is (weakly) worse than attacking at $t = 1$, and it is strictly so regardless of the other player's choices when $\theta = B$, which occurs with $\pi_B > 0$. As a result, waiting at $t = 1$ and then attacking at $t = 2$ regardless of whether $d = 0$ or $d = 1$ (*WAA*) is strictly dominated by attacking right away (*A*). Moreover, recall that the (conditionally) dominant strategy is to attack at $t = 2$ following $d = 1$, which occurs with a strictly positive probability since $\pi_B > 0$. As a result, strategies *WAN* and *WNN* are strictly dominated by *A* and *WNA*, respectively. Therefore, the only strategies that survive the first-round elimination of strictly dominated strategies are attacking right away (*A*) and wait and follow (*WNA*).

In the second round, under the condition $c < \pi_G/(1 - \pi_G)$, strategy *A* is strictly dominated by *WNA*. After the first round of elimination, a rational opponent plays either *A* or *WNA*. Suppose that the opponent plays *A* with probability $q$ and *WNA* with probability $(1 - q)$ for some $q \in [0, 1]$. Then, the expected payoff from playing *WNA* is $2\pi_G + [2(1 - q) + (1 - c)q]\pi_M + (1 - c)\pi_B$, which is strictly decreasing in $q$. This means the expected payoff from *WNA* is at least $2\pi_G + (1 - c)(1 - \pi_G)$, which is achieved when the opponent plays *A* with probability $q = 1$. On the other hand, the payoff from *A* is 1. Therefore, under the condition $c < \pi_G/(1 - \pi_G)$, the expected payoff from wait and follow (*WNA*) is strictly higher than that from attacking right away (*A*) regardless of whether the opponent plays *A* or *WNA*.[5]

Thus, the only strategy that survives two rounds of iterated elimination of strictly dominated strategies is to wait and follow (*WNA*). When both players take this strategy, under $\theta \neq B$, since they will not attack at $t = 1$, there will be no alert in period 2 (i.e., $d = 0$). Following no alert, both players do not attack in period 2. Accordingly, panic is eliminated. This elimination of panic in the above dynamic coordination game requires two conditions: (1) cost—a positive but sufficiently small delay cost $c$—and (2) doubt—$\pi_B$, $\pi_G > 0$. Note that the second condition means that each player always assigns a positive probability that his choice could be a mistake regardless of what the other player does. In this sense, we refer to this condition as *doubt*.[6]

[5] We can interpret $\pi_G \times 1$ as the minimum expected benefit from waiting for the alert—if there is no alert (i.e., $d = 0$), which occurs with a probability of at least $\pi_G$, the player gets 2 rather than 1—and $(1 - \pi_G) \times c$ as the maximum expected cost of waiting—if the alert is triggered (i.e., $d = 1$), which occurs with a probability of at most $1 - \pi_G$, the player gets $1 - c$ rather than 1. This gives us the above inequality.

[6] Note that under the disaster alert, when $\theta = B$, the players attack in period 2 and bear a cost $c$. One may conjecture that to avoid this cost, the principal can supplement a disaster alert with an additional disclosure in period 1 whether or not $\theta = B$. However, this removes doubt ($\pi_B = 0$ after learning $\theta \neq B$), and therefore, our main result does not hold.

### 1. Difference from Static Disclosure

The strategy *WNA* is unique in surviving iterated elimination of strictly dominated strategies. Thus, both players adopting this strategy and the resulting Bayesian consistent beliefs at both $d = 0$ and $d = 1$ (since $\pi_B$, $\pi_G > 0$, both of these information sets are reached with positive probability) constitute the unique sequential equilibrium for the dynamic game induced by the disaster alert policy. Recall that publicly disclosing $\theta \neq B$ before the period 1 play cannot eliminate panic since there is an equilibrium in which both players attack following that information. Interestingly, under the strategy profile where both players take *WNA*, on path, the public information $d = 0$ indeed means $\theta \neq B$.[7] However, there is no sequential equilibrium in which players attack following no alert. To see this, suppose, for contradiction, there is a sequential equilibrium in which the two players attack after seeing no alert ($d = 0$). If one looks at this part of the game in isolation, both players attacking can be an equilibrium of the continuation game. Then, in this sequential equilibrium, at time $t = 1$, a player anticipates that she will attack at $t = 2$ regardless of whether $d = 0$ or $d = 1$ is reached. Recall that it is suboptimal not to attack immediately if one plans to attack regardless of alert or no alert: this goes back to our point on the elimination of the *WAA* strategy. Therefore, one can generate a candidate sequential equilibrium in which both players attack at every information set (including the initial one at $t = 1$). However, since in this candidate equilibrium both players attack immediately, the no alert information set ($d = 0$) is reached if and only if $\theta = G$. But then, on that information set, Bayesian updating requires players to assign probability 1 to $\theta = G$. Under this belief, attacking is a dominated strategy after seeing no alert, which is a contradiction.[8]

Our proposed disclosure policy is simple: a one-time, binary, and public disclosure. It announces at the end of period 1 that "the good outcome is still achievable" (no alert $d = 0$) or not. It is based on both the exogenous $\theta$ and the endogenous period-1 play. Recall that disclosing this information at the beginning cannot eliminate panic. The crucial difference is that the player needs to wait for this information. Since there is no wait involved for the static disclosure, then, under adversarial selection, players may not follow the recommendation after learning "the good outcome is

[7] Different from the disclosure of $\theta \neq B$ before period 1, the message of no alert, by the design of the disaster alert, depends on what strategy players choose in period 1. It means $\theta \neq B$ only when both players choose to wait in period 1.

[8] Our dynamic adversarial information design approach selects the worst-possible outcome for the principal, state by state, across all rationalizable strategies in the strategic-form game. Notably, under this approach, subgames (e.g., starting from $d = 0$) cannot be analyzed in isolation to determine the worst possible outcome within that subgame. We formally define the adversarial information design approach in sec. II.C. We thank an anonymous referee for prompting this clarification.

still achievable." However, as long as each player has to pay a delay cost to wait, then, if the player chooses to wait, based on the strict dominance, he would follow the recommendation: attack if and only if "the good outcome is not possible." Moreover, if the cost of waiting is sufficiently small and the players have doubt, they will choose to wait.

## 2. Forward Induction

This result may remind readers of the forward induction argument, pioneered by Kohlberg and Mertens (1986). To see this connection, let us simplify the example further. Suppose that the state is medium (no uncertainty about the state) and only player 1 (not both players) has the option to attack early. In this case, both players attacking at the first opportunity they get is an equilibrium, while both waiting and then coordinating on not attacking is another equilibrium. Intuitively, the forward induction argument suggests that one player can signal his intention of not attacking later to the other player by not taking the outside option to attack early. Formally, iterated elimination of weakly dominated strategies eliminates the former equilibrium and selects the latter one (see sec. I.1 of the online appendix for the formal result).[9]

In the applications we have in mind, the natural assumption is that all players have the option to attack early. However, when both players can attack early, this forward induction argument does not hold. Intuitively, if a player believes that the other player will not wait, then he has no incentive to signal his intention. Formally, iterated elimination of weakly dominated strategies does not eliminate the equilibrium where both players attack at period 1 (see sec. I.2 of the online appendix). From this comparison, it should be clear that fundamental uncertainty is critical to our result. Our disclosure policy exploits this uncertainty about the state (the no alert information set is reached with a probability of at least $\pi_G > 0$), inducing the players to wait and follow even when they believe that the other player may attack right away. Thus, our result is similar to forward induction in selecting the efficient equilibrium in a coordination game, but the mechanism differs. Moreover, the fundamental uncertainty allows us to use iterated strict dominance instead of weak dominance as our solution concept.

## 3. Beyond the Simple Example

We show that the insights derived from our simple example can be applied to a fairly general canonical regime change setting, which accommodates flexible payoff specifications and information structures. In this general

[9] Ben-Porath and Dekel (1992) have shown the equivalence of forward induction and iterated weak dominance.

model with a continuum of agents and continuous states, each agent makes decisions on when to attack within a time window $[0, T]$ (if the agent chooses to attack at all). The distinct advantage of this continuous-time setting lies in the principal's ability to select the timing of disclosure, which makes the condition $c < \pi_G/(1 - \pi_G)$ (as in the simple example) natural when the duration of waiting is short. Our general setup has two main ingredients. First, an agent incurs a cost from delayed attacks, and this cost is bounded above by a function $c(t)$ that is continuously increasing in the duration of waiting ($t$) and $c(0) = 0$. Second, we posit the existence of a positive lower bound, denoted as $\epsilon$, on each agent's belief regarding the state lying within dominance regions. Essentially, this assumption implies that each agent holds the belief that initiating or refraining from an attack is a mistake (regardless of others' actions) with a probability of at least $\epsilon$. We refer to this assumption as *doubt*.

Under these two assumptions—cost and doubt—we establish the existence of $\hat{\tau} > 0$, such that, given a disaster alert scheduled at any time $\tau \in (0, \hat{\tau})$, the unique rationalizable strategy for any agent of any type is to wait and follow the alert, thereby eliminating panic even under adversarial selection. Intuitively, our disaster alert is a message from the principal: "Do not panic, just wait a little (until time $\tau$), and if a disaster becomes inevitable, I will let you know."

Let us first consider the assumption regarding cost. As we can see from our simple example, both players may attack immediately rather than wait for the disaster alert if $c$ is sufficiently high. The condition delay cost may not apply universally across all dynamic coordination games. However, in section IV, we provide three classic examples of a regime change game demonstrating why this cost condition holds naturally under a continuous-time setup. For instance, in the case of sovereign debt default, this property holds because the sovereign debt default, if it occurs at all, occurs only at a fixed date when the debt matures. Therefore, a short wait does not involve much cost. In contrast, in the dynamic bank run application, a bank can fail instantaneously as a result of endogenous attacks over time, and attacking earlier than the occurrence of disaster yields a significant first-mover advantage. Nevertheless, if the bank becomes vulnerable only after a negative shock arrives, then the cost of waiting for a short duration should be small since the probability of the negative shock arriving during a brief waiting period is low. Additionally, in cases of "slow-moving" capital where the full execution of an attack takes time (e.g., firing employees or selling properties takes time), even though a disaster might occur at any moment, if only a limited portion of the attack can be completed before the disaster because of short waiting periods, the cost of waiting can effectively be made negligible under a timely disaster alert.

Next, we turn to the doubt assumption. We can see from our simple example that if $\pi_G = 0$, then both players attacking immediately constitutes

an equilibrium. Therefore, this assumption is necessary when the agents share the same beliefs. Our main result shows that under the doubt assumption, the same intuition applies even when players receive arbitrarily correlated private signals. Interestingly, when the agents receive private signals, a disaster alert can be effective even when some types do not assign a positive probability to $\theta = G$. An agent who receives such a signal may assign a positive probability that his opponent has doubt, and so, his opponent will wait for the disaster alert. Then, he can be persuaded to wait for the disaster alert as well even though he assigns a zero probability to $\theta = G$. In section V.A, we illustrate this using our simple example. This argument can be extended to our general setup as well by "infecting" more types who believe that the others may not have doubt either, but others believe that others have doubt, and so on. In section V.B, we specialize to a global game information structure with independent Gaussian signals (which violates the doubt assumption) and show that disaster alert eliminates panic in the limit. For simplicity of exposition, we maintain the assumption that the agents do not observe any new information over time. In section V.C, we show that the result is even stronger when the agents can observe whether the disaster happens. In section V.D, we introduce the arrival of exogenous information over time. We show that if such information arrives somewhat infrequently, the designer can eliminate panic by setting timely disaster alerts every time after information arrives and before the next time the exogenous information arrives.

We conclude by highlighting some of the limitations of early warnings. We point out that in our dynamic coordination game, the agents can choose when to attack, and the attack is the only irreversible choice. Moreover, our principal has a simple objective: she wants to avoid a disaster. We argue that an early warning may not be as effective (if not completely ineffective) in the absence of these features.

*Related literature.*—Theoretically, this paper contributes to two strands of the literature: (1) dynamic coordination and (2) Bayesian persuasion or information design. See Angeletos and Lian (2017) for a recent survey on dynamic coordination. This literature studies how coordination can be affected when agents learn some information about the past (see, e.g., Chamley 1999, 2003; Angeletos, Hellwig, and Pavan 2007; Dasgupta 2007; Chassang 2010). Unlike the abovementioned papers, we study how a principal can manipulate the agents' beliefs by optimally choosing when and how the agents learn about past attacks and the underlying state.

There is a substantial and growing body of work on Bayesian persuasion or information design, which started with Kamenica and Gentzkow (2011). See Bergemann and Morris (2019) and Kamenica (2019) for recent surveys. We adopt the adversarial approach to study dynamic information design in a dynamic coordination game, which often is associated with strategic uncertainty and gives rise to multiple equilibria. In our setup, if

the principal has the power to select which equilibrium the agents will play (as in partial implementation), then, consistent with the revelation principle, she can simply recommend "attack" when the fundamental warrants a run and "never attack" otherwise. In this way, the principal attains her desired outcome. However, note that there can be a bad equilibrium in which the agents are not obedient. The revelation principle does not hold under full implementation. Mathevet, Perego, and Taneva (2020) have shown that when the principal can only choose the information structure, whereas the agents play the worst equilibrium (adversarial selection), the revelation principal argument breaks down. Under adversarial selection, Goldstein and Huang (2016), Li, Song, and Zhao (2023), and Inostroza and Pavan (2025) analyze the optimal information design in a static regime change game. The authors find that the optimal design achieves perfect coordination, but the principal fails to achieve her desired outcome under adversarial selection. For details, please refer to section III.A. In contrast, we propose a simple two-stage public disclosure through which the principal achieves her desired outcome based on the unique rationalizable strategy profile.[10]

It is worth mentioning that some recent studies have looked into Bayesian persuasion in dynamic settings (see, e.g., Au 2015; Ely 2017; Ely and Szydlowski 2020; Orlov, Skrzypacz, and Zryumov 2020; Ball 2023). Broadly speaking, this literature uses future disclosure as a carrot to persuade an agent to do what the principal wants. A few studies have also addressed the optimal information design in some dynamic games. For instance, Li, Szydlowski, and Yu (2025) find the optimal disclosure in a dynamic entry game, and Ely et al. (2023) find the optimal disclosure in a dynamic contest. We study the optimal information design in a dynamic coordination game. Our main contribution is to show how to exploit the time dimension to persuade agents who play the adversarial equilibrium in a coordination game.

The rest of the paper is organized as follows. Section II describes the model and solution concept. Section III shows how a timely disaster alert eliminates panic under adversarial selection. Section IV discusses three applications of our theory and, in particular, how the condition on delay

[10] The two-stage mechanism can be interpreted as follows: (1) in the first stage, it recommends that everyone "not attack and wait" for the second-stage recommendation; (2) in the second stage, it recommends (to all agents who wait) that they "attack" if, based on $\theta$ and the endogenous attack (who did not wait), a disaster (regime change) is inevitable and "not attack" otherwise. We identify the conditions of exogenous information structure and the cost of delay under which the unique rationalizable strategy is to follow the two-stage recommendations for all possible types.

While we consider a specific objective the principal wants to implement, it is worth mentioning that a mechanism design literature, following Moore and Repullo (1988), constructs multistage mechanisms to fully implement social choice functions that cannot be implemented by static mechanisms as a result of violating Maskin monotonicity.

cost gets satisfied in those applications. Section V considers to what extent the restrictions on the exogenous information structure can be relaxed, and section VI concludes by discussing some limitations of early warnings.

## II. Model

### A. *Dynamic Coordination Game*

The economy is populated by a principal (she) and a continuum of risk-neutral agents (he), indexed by $i \in [0, 1]$. Before the game begins, nature draws a state $\theta \in \Theta \subseteq \mathbb{R}$ from a commonly known distribution $\Psi : \Theta \to [0, 1]$. An agent $i$ receives a noisy signal $s_i \in \mathbb{S}$. Given any underlying fundamental state $\theta$, the signal profile $s(\theta) \in \mathbb{S}^{[0,1]}$ is drawn from a joint distribution $F(s|\theta)$ with associated density $f(s|\theta)$. An agent $i$ chooses when to attack, $t_i \in [0, T]$, if at all. We represent not attacking in the given time window as $t_i = \infty$. We define the mass of attack until time $t$ (inclusive) as $N(t) \coloneqq \int_i \mathbb{1}(t_i \leq t)\mathrm{d}i$, which is nondecreasing in time $t$, the time path of aggregate attack as $\mathbf{N} \coloneqq (N(t))_{t\in[0,T]}$, and the set of the possible time path of aggregate attack as $\mathbb{N} = [0, 1]^{[0,T]}$. The time path of aggregate attack until time $t$ (not inclusive) is defined as $\mathbf{N}_t = (N(t'))_{t'\in[0,t)}$ and the set of all possible continuation time paths that follow $\mathbf{N}_t$ as $\mathbb{N}|\mathbf{N}_t \coloneqq \{\hat{\mathbf{N}} \in \mathbb{N} | \hat{N}(t') = N(t') \text{ for } t' \in [0, t)\}$. Let $u(t_i, \theta, \mathbf{N})$ be the payoff of an agent $i$ who plays $t_i$ while the state is $\theta$ and the time path of aggregate attack is $\mathbf{N}$.

*Disaster.*—We say that the game ends in a *disaster* if and only if $(\theta, \mathbf{N}) \in \mathcal{D}$, for some $\mathcal{D} \subset \Theta \times \mathbb{N}$. Everything else the same, a disaster is more likely to occur if (1) the fundamental $\theta$ is smaller, (2) some agents switch from not attacking (i.e., $t_i = \infty$) to attacking (i.e., $t_i \in [0, T]$), or (3) some agents choose to attack at an earlier date (i.e., $t_i$ decreases). Formally, the event disaster $\mathcal{D}$ satisfies the following properties:

(1) if $(\theta, \mathbf{N}) \in \mathcal{D}$, then $\forall\, \theta' \leq \theta$, $(\theta', \mathbf{N}) \in \mathcal{D}$; and
(2) if $(\theta, \mathbf{N}) \in \mathcal{D}$, then $\forall\, \hat{\mathbf{N}} \geq \mathbf{N}$, $(\theta, \hat{\mathbf{N}}) \in \mathcal{D}$,

where $\hat{\mathbf{N}} \geq \mathbf{N}$ means $\hat{N}(t) \geq N(t)$ for all $t \in [0, T]$. For any given time path of aggregate attack until time $t$, $\mathbf{N}_t$, a possible continuation time path of aggregate attack in $\mathbb{N}|\mathbf{N}_t$ is one in which no one attacks following $\mathbf{N}_t$ from time $t$ to the end. This particular time path of aggregate attack is denoted by $\mathbf{N}^0|\mathbf{N}_t$. By definition of disaster, if $(\theta, \mathbf{N}^0|\mathbf{N}_t) \in \mathcal{D}$, then for all $\hat{\mathbf{N}} \in \mathbb{N}|\mathbf{N}_t$, $(\theta, \hat{\mathbf{N}}) \in \mathcal{D}$ since $\hat{\mathbf{N}} \geq \mathbf{N}^0|\mathbf{N}_t$. In other words, if at time $t$, $(\theta, \mathbf{N}^0|\mathbf{N}_t) \in \mathcal{D}$, then a disaster is inevitable regardless of what agents do thereafter.

We say that the fundamental is in the lower (upper) dominance region $\Theta^L$ ($\Theta^U$), if, regardless of the agents' action, a disaster is inevitable

(impossible). We assume lower and upper dominance regions are not empty; that is, $\Theta^L := \{\theta \in \Theta | \ \forall\ \mathbf{N} \in \mathbb{N}, (\theta, \mathbf{N}) \in \mathcal{D}\} \neq \varnothing$ and $\Theta^U := \{\theta \in \Theta | \ \forall\ \mathbf{N} \in \mathbb{N}, (\theta, \mathbf{N}) \notin \mathcal{D}\} \neq \varnothing$.[11]

*Complementarity.*—In a canonical static regime change game, agents strictly prefer attacking (not attacking) if a disaster (no disaster) occurs. This property naturally extends to our dynamic setting for attacking at any date $t \in [0, T]$; that is,

$$\forall\ t \in [0, T] \text{ and } (\theta, \mathbf{N}) \in \mathcal{D}, \quad u(\infty, \theta, \mathbf{N}) < u(t, \theta, \mathbf{N}); \text{ and} \tag{1}$$

$$\forall\ t \in [0, T] \text{ and } (\theta, \mathbf{N}) \notin \mathcal{D}, \quad u(\infty, \theta, \mathbf{N}) > u(t, \theta, \mathbf{N}). \tag{2}$$

This payoff specification exhibits complementarity in the following sense. Recall that by definition of a disaster, if others are more likely to attack, or they attack at earlier dates (for a given aggregate attack $N(T)$), then the game is more likely to end in a disaster, and therefore an agent would have a greater incentive to attack.

We impose the following restriction on the general payoff specification. The benefit of not attacking over attacking in the event of no disaster has a strictly positive lower bound $g > 0$:[12]

$$u(\infty, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) \geq g, \quad \forall\ t \in [0, T], (\theta, \mathbf{N}) \notin \mathcal{D}. \tag{3}$$

*Cost of delayed attack.*—Different from the static regime change game, the agents can choose the timing of the attack. The following assumption governs the agent's preference over the attack time $t \in [0, T]$.

Assumption 1 (Costly delay). The dependence of $u(t, \theta, \mathbf{N})$ on $t$ satisfies that

(*A*) it is decreasing in $t \in [0, T]$ for any $(\theta, \mathbf{N})$, and it is strictly decreasing if $\theta \in \Theta^L$;

(*B*) there exists a continuous and strictly increasing function $c(t)$ defined on $t \in [0, T]$ with $c(0) = 0$, such that for any $(\theta, \mathbf{N}) \in \mathcal{D}$ and $t \in (0, T]$,

$$u(0, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) \leq c(t). \tag{4}$$

Assumption 1(A) states that the delayed attack is costly and strictly so only when $\theta \in \Theta^L$. As agents always assign a strictly positive probability to

[11] For example, suppose that $\Theta = \mathbb{R}$ and the game ends in a disaster ($(\theta, \mathbf{N}) \in \mathcal{D}$) when the fundamental is not strong enough to sustain the aggregate attack until time $T$, i.e., $\theta - N(T) \leq 0$. Notice that it satisfies the above two properties. The lower and upper dominance regions in this case are $(-\infty, 0]$ and $(1, \infty)$, respectively.

[12] For standard applications of regime change (see sec. IV), these payoff differences are assumed to be constants, which means this assumption trivially holds. Here, we allow for more general payoff specifications.

the states $\theta \in \Theta^L$ (under assumption 2[A] introduced later), this condition is sufficient to ensure that any delay in attack is strictly costly (in expectation). Therefore, if an agent prefers attacking (e.g., when the game ends in a disaster), he would attack as soon as he can. In addition, assumption 1(B) requires, whenever the game ends in a disaster, the cost of delay to be bounded above by a function $c(\cdot)$ that is continuously increasing in the duration of waiting. This assumption is a necessary ingredient for our main result. In section IV, we provide three natural applications of dynamic regime change that satisfy this assumption.

*Exogenous information.*—The signals are drawn from a known joint distribution $F(s|\theta)$. The signals can be arbitrarily correlated. This allows for conditionally independent signals (as is standard in global games) as well as perfectly correlated signals leading to homogeneous beliefs (as in our example in the introduction). We only impose the following restriction on the exogenous information structure.

Assumption 2 (Doubt). Any agent $i$ with any signal $s_i \in \mathbb{S}$ believes that[13]

(*A*) $\mathbb{P}(\theta \in \Theta^L | s_i) > 0$, and
(*B*) there exists some $\epsilon > 0$ such that, for any $s_i \in \mathbb{S}$, $\mathbb{P}(\theta \in \Theta^U | s_i) > \epsilon$.

This assumption holds true if, for instance, the marginal distribution of the signal has full support and a bounded density. Formally, if $f_i(s_i|\theta) \in [\underline{f}, \bar{f}]$ for all $\theta \in \Theta$, and $s_i \in \mathbb{S}$, where $0 < \underline{f} \leq \bar{f} < +\infty$, then assumption 2 holds true for any $\epsilon \in (0, \int_{\theta \in \Theta^U} d\Psi(\theta) \underline{f}/\bar{f})$.

Assumption 2(A) guarantees that regardless of the private signal an agent receives, he assigns a positive probability that the game will end in a disaster. Therefore, regardless of others' strategies, he has doubt whether not attacking is the correct action. Conversely, assumption 2(B) implies that regardless of his signal and others' strategies, he has doubt whether attacking is the correct action.

Note that assumption 2(B) is more demanding. It also assumes this doubt has a strictly positive lower bound $\epsilon$. In section V, we show that assumption 2(B) can be relaxed to some extent when agents have heterogeneous beliefs. In section V.A, we revisit the example from the introduction and show that the same result holds even though some types may

[13] For any $\hat{\Theta} \subseteq \Theta$,

$$\mathbb{P}(\theta \in \hat{\Theta} | s_i) = \frac{\int_{\theta \in \hat{\Theta}} f_i(s_i|\theta) d\Psi(\theta)}{\int_{\theta \in \Theta} f_i(s_i|\theta) d\Psi(\theta)},$$

where $f_i(s_i|\theta) = \text{marg}_{s_{-i}} f(s|\theta)$.

assign zero probability to $\theta \in \Theta^U$. In section V.B, we show the argument can be extended to a standard global game setup with independent Gaussian noise, which violates assumption 2(B).[14] For simplicity of exposition, we assume that the agents do not observe any new information over time. We relax this assumption in sections V.C and V.D and show how our main result can be extended.

### B. *Information Design*

The principal's payoff is $\mathbb{1}((\theta, \mathbf{N}) \notin \mathcal{D})$; that is, she gets 0 if the game ends in a disaster and 1 otherwise. The principal is able to set a disclosure policy based on the fundamental state $\theta$ although she has no control over the realization of $\theta$. The principal does not know the private signals that the agents have received. However, unlike the agents, at any time $t$, the principal can observe the endogenous history of attacks until then; that is, $\mathbf{N}_t = (N(t'))_{t'<t}$. Let $\mathcal{F}_t$ be all information available to the principal by time $t$ and $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$. Before time 0, the principal commits to a dynamic public information disclosure policy denoted by $\Gamma = (d, \mathcal{S})$, where $\mathcal{S}$ is the message space, and $d$ maps $\mathbb{F}$ to $\Delta(\mathcal{S})$.

### C. *Solution Concept*

We consider the strategic form representation of the dynamic game. Given the disclosure policy $\Gamma$, a strategy of an agent of any type is to make a contingency plan for when to attack ($t_i \in [0, T] \cup \{\infty\}$) at the initial history $\varnothing$ and at every history based on the information disclosed by the principal. We use iterated elimination of strictly dominated strategies in the strategic form game as our solution concept. A strategy is rationalizable if it can survive the iterated elimination of strictly dominated strategies. Given any disclosure policy $\Gamma$, let $\mathcal{R}(\Gamma)$ be the set of rationalizable strategy profiles.[15]

The principal does not expect the agents to play the rationalizable strategy profile that is advantageous to her. Rather, she anticipates, state by state, the worst-case scenario that is consistent with some rationalizable strategy profile. We call this *adversarial selection*. Define[16]

$$\Theta^{\mathcal{D}}(\Gamma) := \{\theta \in \Theta | (\theta, \mathbf{N}) \in \mathcal{D} \quad \text{for some} \quad \chi \in \mathcal{R}(\Gamma) \text{ that induces } \mathbf{N}\}.$$

[14] This is because, for any $\epsilon$, there exists a sufficiently small signal that violates the inequality in assumption 2(B).

[15] Note that the set of rationalizable strategy $\mathcal{R}(\Gamma)$ could be different based on the underlying exogenous information structure $F$. However, since $F$ is exogenous and commonly known, for convenience, we suppress $F$ from $\mathcal{R}(\cdot)$.

[16] It should be noted that $\mathbf{N}$, under strategy profile $\chi$ and policy $\Gamma$, becomes some measurable function of the realized fundamental $\theta$.

In words, if $\theta \notin \Theta^{\mathcal{D}}(\Gamma)$, then the game will not end in a disaster, regardless of whatever rationalizable strategies the agents play; otherwise, if $\theta \in \Theta^{\mathcal{D}}(\Gamma)$, the game may end in a disaster under some rationalizable strategy profile $\chi \in \mathcal{R}(\Gamma)$.

Recall that when $\theta \in \Theta^{L}$, the game will inevitably end in a disaster, regardless of what the agents do. Hence, $\Theta^{L} \subseteq \Theta^{\mathcal{D}}(\Gamma)$ for any $\Gamma$. However, the game can end in a disaster even when it is not warranted ($\theta \notin \Theta^{L}$), as agents may choose to attack if they believe that the disaster may occur as a result of others' attacks.

DEFINITION 1 (Panic). For any disclosure policy $\Gamma$, panic is the event where the fundamental state $\theta \in \Theta^{P}(\Gamma) := \Theta^{\mathcal{D}}(\Gamma) - \Theta^{L}$.

This means under the policy $\Gamma$, when $\theta \in \Theta^{P}(\Gamma)$, the game may end in a disaster under some rationalizable strategy profile, and if it happens, it is caused by panic-based attacks. Given the principal's payoff specification and under adversarial selection, she chooses $\Gamma$ to maximize $\mathbb{E}[\mathbb{1}(\theta \notin \Theta^{\mathcal{D}}(\Gamma))]$. Since a disaster is unavoidable regardless of $\Gamma$ for any $\theta \in \Theta^{L}$, the principal's objective is equivalent to minimizing the ex ante probability of panic; that is,

$$\min_{\Gamma} \ \mathbb{P}(\theta \in \Theta^{P}(\Gamma)),$$

where $\mathbb{P}(\theta \in \Theta^{P}(\Gamma)) = \int_{\Theta^{P}(\Gamma)} d\Psi(\theta)$. We refer to this optimization as *optimal dynamic adversarial information design*. If $\Theta^{P}(\Gamma) = \varnothing$, then we say that policy $\Gamma$ eliminates panic in a robust sense.

## III. Main Result

In this section, we construct a simple dynamic information disclosure policy. This policy induces (even in the worst case) the agents to perfectly coordinate their actions and never attack a regime when it is not warranted ($\theta \notin \Theta^{L}$), thereby eliminating panic in the robust sense.

### A. *Static Benchmarks*

To fix ideas, we start with no disclosure and then consider some simple time-0 disclosure policies, where the principal sends the message only about the fundamental $\theta$ before the agents have their first opportunity to attack. We show that these history-independent disclosure policies cannot eliminate panics in the robust sense.

*No disclosure.*—Suppose that the principal does not provide any additional information. Since the delayed attack is costly and strictly so when $\theta \in \Theta^{L}$ (assumption 1[A]), whereas an agent always assigns a positive probability that $\theta \in \Theta^{L}$ regardless of the signal $s_i$ (assumption 2[A]), if an agent decides to attack, he will do so immediately at time 0. Thus,

the game boils down to the canonical static regime change game where the agents choose between attacking right away or never attacking. Depending on the distribution of the signals that the agents receive, there can be multiple equilibria. For instance, suppose that the agents share a homogeneous belief about $\theta$. If they believe that the probability that $\theta \in \Theta^L$ is sufficiently small, there exists an equilibrium in which they do not attack. However, if they believe that the probability that $\theta \in \Theta^U$ is sufficiently small, then there also exists another equilibrium in which they all panic and attack. Using global game perturbation, Morris and Shin (2003) show that if the agents have sufficiently heterogeneous beliefs, there is a unique rationalizable strategy: attack if and only if the private signal is below a threshold. This means that unless the fundamental is sufficiently strong, attacks may be based on panic and a disaster may occur as a result. (see sec. V.B).

*Time-0 full disclosure.*—Suppose that the fundamental $\theta$ is fully revealed at time 0. A possible equilibrium outcome is that all the agents perfectly coordinate on not attacking when $\theta \notin \Theta^L$. However, given the strategic complementarity, this is not the only rationalizable strategy profile. In another possible equilibrium, all agents attack at $t = 0$ whenever $\theta \notin \Theta^U$. Therefore, for a principal who anticipates the adversarial outcome, full disclosure is the worst policy because it maximizes the chance of disaster.

*Time-0 partial disclosure.*—Many time-0 partial disclosure policies exist. Consider the following policy (denoted as $\Gamma^0$): at time 0, the principal publicly discloses whether or not the regime change is warranted ($\theta \in \Theta^L$). Recall that since a delayed attack is strictly costly if an agent attacks, he attacks at time 0. In one equilibrium, the agents attack at time 0 if the principal sends the message that the regime change is warranted; otherwise, they do not attack. However, panic-based attacks are possible in other equilibria (see Angeletos, Hellwig, and Pavan 2007). To ensure that the agents never attack in any equilibrium, the positive news needs to be stronger than $\theta \notin \Theta^L$. Under conditionally independent signals, Goldstein and Huang (2016) find the optimal monotone public disclosure policy. The authors show that if the principal discloses whether $\theta > k$ for a sufficiently large $k$, then the agents will never attack when $\theta > k$. Inostroza and Pavan (2025) show that this may not be the optimal disclosure policy since nonmonotone disclosure can sometimes yield a better result. The authors investigate when monotone public disclosure is indeed optimal. The public disclosure could be suboptimal. Li, Song, and Zhao (2023) consider a static regime change game in which the principal is the only source of information; that is, the agents have no private information. They study the optimal way to provide private information to the agents to save as many states as possible. They find an optimal design where the principal locally obfuscates, that is, sends a signal matching the regime's true strength to some and an elevated signal professing slightly higher

strength to others. Morris, Oyama, and Takahashi (2024) consider a more general binary action supermodular game setup. They show that the optimal design always involves perfect coordination if the payoffs are not too asymmetric. Nevertheless, the principal fails to attain her most preferred outcome.

Note that these studies consider variations of a static regime change game and provide some interesting insights. However, under all these optimal policies, the game may still end in a disaster even though the fundamental does not warrant it ($\theta \notin \Theta^L$). Therefore, although a time-0 disclosure policy can improve the worst possible outcome, panics cannot be avoided in the robust sense.

The above discussion points out two important aspects of the problem. First, to understand how to reduce the chance of panic, we cannot simply select the principal's preferred equilibrium. Many policies can eliminate panic if the agents play the principal's preferred equilibrium (e.g., full disclosure, or $\Gamma^0$ policy). However, these policies cannot eliminate panic in the robust sense. Second, panic cannot be eliminated simply by providing information early (at time 0, before the agents have the chance to attack). Below, we show that the principal can exploit the endogenous waiting and construct a simple dynamic disclosure policy that eliminates panic in the robust sense.

### *B. A Simple Dynamic Disclosure Policy: Disaster Alert*

We consider a simple dynamic partial disclosure policy. The principal sends a binary public signal at a fixed date $\tau > 0$, based on the fundamental state $\theta$ and the time path of aggregate attack until time $\tau$ (not inclusive), $\mathbf{N}_\tau = (N(t))_{t\in[0,\tau)}$. Note that $\mathbf{N}_\tau$ is determined by strategy profile $\chi$ the agents endogenously choose to play, which in turn depends on the disclosure policy.

Definition 2 (Disaster alert). A disaster alert, denoted by $\Gamma^\tau$, is a public information disclosure at time $\tau \in (0, T)$ where the message space is $\mathcal{S} = \{0, 1\}$ and disclosure rule is

$$d^\tau(\theta, \mathbf{N}_\tau) = \begin{cases} 1 & \text{if } (\theta, \mathbf{N}^0|\mathbf{N}_\tau) \in \mathcal{D}, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that $\mathbf{N}^0|\mathbf{N}_\tau$ is the continuation time path of aggregate attack in which no agent attacks following $\mathbf{N}_\tau$. We say that the alert is triggered if $d^\tau = 1$ and is not triggered if $d^\tau = 0$. Note that if the alert is triggered—that is, $\theta$ and $\mathbf{N}_\tau$ are such that $(\theta, \mathbf{N}^0|\mathbf{N}_\tau) \in \mathcal{D}$—then by definition of disaster, $(\theta, \hat{\mathbf{N}}) \in \mathcal{D}$ for all $\hat{\mathbf{N}} \in \mathbb{N}|\mathbf{N}_\tau$. This means regardless of the continuation time path of aggregate attack following $\mathbf{N}_\tau$, a disaster is inevitable. On the other hand, if the alert is not triggered, it implies that

the game will not end in a disaster if the agents who have waited for the disclosure do not attack after learning $d^\tau = 0$. However, if some agents choose to attack after no alert, the game may end in a disaster. That is, for any $\theta$ and $\mathbf{N}_\tau$ such that $d^\tau(\theta, \mathbf{N}_\tau) = 0$, it is possible to have some $\hat{\mathbf{N}} \in \mathbb{N} | \mathbf{N}_\tau$ such that $(\theta, \hat{\mathbf{N}}) \in \mathcal{D}$. The idea behind this disaster alert is that, by triggering the alert, the principal informs the agents that attacking is now the dominant choice since the game will surely end in a disaster regardless of what the agents do after disclosure.

In this spirit, the proposed disaster alert policy resembles a real-world early warning system, which aims to generate an accurate forecast of a future crisis by all historical information. Notice that the proposed policy is simple: it is binary and public. However, it is also history dependent and thus an *endogenous disclosure*. The message disclosed at date $\tau$ depends on the history of attack before the time of disclosure, $\mathbf{N}_\tau$. If some agents do not wait for the disaster alert and attack before time $\tau$, then their actions may trigger the alert.

### C. *Optimality of Timely Disaster Alert*

The proposed policy can be interpreted simply as assurance from the principal: "Do not panic; just wait until time $\tau$, and at that time, I will send you an alert if a disaster is inevitable." The following theorem shows that if the disaster alert is scheduled to be disclosed in a sufficiently short period of time, then under the unique rationalizable strategy, the agents will always wait and follow the principal's advice regardless of their own information. According to this unique rationalizable strategy profile, the game ends in a disaster only when $\theta \in \Theta^L$. In other words, a timely disaster alert eliminates panic in the robust sense.

THEOREM 1. Under assumptions 1 and 2, for any $\tau \in (0, \hat{\tau})$ where $\hat{\tau} = \min\{c^{-1}(g \times \epsilon / (1 - \epsilon)), T\} > 0$, under the disclosure policy $\Gamma^\tau$, panic is eliminated in the robust sense; that is, $\Theta^P(\Gamma^\tau) = \varnothing$.

Theorem 1 claims that when the principal sets the disaster alert in a timely manner, she achieves her most preferred outcome as the unique rationalizable outcome. We prove this theorem using three lemmas that we will discuss next. It will be useful to formalize the strategy and the expected payoffs in the dynamic coordination game induced by the disaster alert policy $\Gamma^\tau$ for $\tau \in (0, T)$. Nature moves first and selects $(\theta, s)$, where $s$ is the profile of signals. Each agent $i \in [0, 1]$ learns his own type $s_i$. Under any disaster alert policy $\Gamma^\tau$, an agent $i$ of type $s_i$ makes a contingency plan for when to attack at three information sets: the initial history ($\varnothing$), the one after the alert is not triggered ($d^\tau = 0$), and the one after the alert is triggered ($d^\tau = 1$). Let $t_\varnothing$, $t_0$, and $t_1$ be the time of attack at these three information sets, respectively. Therefore, the space of contingency plans in the dynamic game induced by the disaster alert $\Gamma^\tau$ is

$$\mathcal{X} \coloneqq \{(t_\varnothing, t_0, t_1) \in ([0, T] \cup \{\infty\}) \times ([\tau, T] \cup \{\infty\})^2\}.$$

In the strategic form game, a strategy $\chi_i$ is a mapping from type space $\mathcal{S}$ to $\mathcal{X}$. Suppose that agent $i$ of type $s_i$ chooses a plan $\chi_i(s_i) = (t_\varnothing, t_0, t_1) \in \mathcal{X}$. Under this plan, the realized time of attack, denoted as $\mathrm{t}_i$, is as follows:

$$\mathrm{t}_i(t_\varnothing, t_0, t_1) = \begin{cases} t_\varnothing \text{ if } t_\varnothing \in [0, \tau); \\ t_0 \text{ if } t_\varnothing \geq \tau \text{ and } d^\tau(\theta, \mathbf{N}_\tau) = 0; \\ t_1 \text{ if } t_\varnothing \geq \tau \text{ and } d^\tau(\theta, \mathbf{N}_\tau) = 1. \end{cases}$$

If $\mathrm{t}_i = \infty$, then we say no attack is realized. Let $\mathbb{U}((t_\varnothing, t_0, t_1), \chi_{-i}; s_i)$ be the expected payoff of agent $i$ of type $s_i$ from playing $(t_\varnothing, t_0, t_1)$ when other agents play strategy $\chi_{-i}$. If $t_\varnothing \in [0, \tau)$, then

$$\mathbb{U}((t_\varnothing, t_0, t_1), \chi_{-i}; s_i) = \mathbb{E}[u(t_\varnothing, \theta, \mathbf{N}(\chi_{-i}))|s_i],$$

and if $t_\varnothing \in [\tau, T] \cup \{\infty\}$, then

$$\begin{aligned} &\mathbb{U}((t_\varnothing, t_0, t_1), \chi_{-i}; s_i) \\ &= \mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 0|s_i)\mathbb{E}[u(t_0, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 0] \\ &+ \mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 1|s_i)\mathbb{E}[u(t_1, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 1]. \end{aligned}$$

Step 1: Option Value of Waiting

LEMMA 1. Under assumptions 1(A) and 2(A), for any agent $i$, the only strategies that survive the first round of elimination of strictly dominated strategies are such that for any given $s_i \in \mathbb{S}$, $\chi_i(s_i) \in \{\mathcal{A}, \mathcal{W}\}$, where

(1) $\mathcal{A}$ means attack immediately: $(0, t_0, t_1)$, where $t_0, t_1 \in [\tau, T] \cup \{\infty\}$; and
(2) $\mathcal{W}$ means wait and follow: $(t_\varnothing, \infty, \tau)$, where $t_\varnothing \in [\tau, T] \cup \{\infty\}$.

*Proof.* We prove this lemma using three simple steps.[17] First, for any agent with signal $s_i$, any plan that involves $t_\varnothing \in (0, \tau)$ is strictly dominated by attacking immediately at $t_\varnothing = 0$ (or $\mathcal{A}$). Formally, for all $s_i$,

$$(0, t_0, t_1) \succ (t_\varnothing, t_0, t_1) \text{ for all } t_\varnothing \in (0, \tau), t_0, t_1 \in [\tau, T] \cup \{\infty\}.$$

To see this, note that, regardless of $s_i$ and $\chi_{-i}$, if an agent does not wait until the disclosure date but attacks at $t_\varnothing \in (0, \tau)$, then he gets a strictly

[17] We have omitted some details in the proof here for brevity and relegated them to the appendix.

lower payoff than attacking at $t_\varnothing = 0$. The payoff is strictly lower because the delayed attack is costly and strictly so when $\theta \in \Theta^L$ (assumption 1[A]), whereas an agent always assigns a positive probability that $\theta \in \Theta^L$ regardless of the signal $s_i$ (assumption 2[A]).

Second, for any agent with any signal $s_i$, all strategies that involve waiting until the disclosure date but not attacking right away following $d^\tau = 1$ are strictly dominated by a plan that involves waiting until the disclosure date and attacking right away at $\tau$ conditional on $d^\tau = 1$. Formally, for all $s_i$,

$$(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, t_1) \text{ for all } t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}, \text{ and } t_1 \in (\tau, T] \cup \{\infty\}.$$

To see this, note that the disaster alert can be triggered because of fundamental reason ($\theta \in \Theta^L$) or because of the endogenous attacks before $\tau$. In either case, a disaster is inevitable following $d^\tau = 1$; that is, $(\theta, \hat{\mathbf{N}}) \in \mathcal{D}$ for all $\hat{\mathbf{N}} \in \mathbb{N}|\mathbf{N}_\tau$. Therefore, for any $\theta$ and $\hat{\mathbf{N}} \in \mathbb{N}|\mathbf{N}_\tau$, it follows from complementarity (see [1]) that when the alert is triggered, an agent strictly prefers attacking at $t_1 = \tau$ over not attacking ($t_1 = \infty$). Moreover, regardless of $s_i$, an agent $i$ assigns a strictly positive probability to the event that $\theta \in \Theta^L$ (assumption 2[A]) and, hence, to the event of an alert being triggered. Therefore, for all $s_i$, and for all $t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}$, $(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, \infty)$. Moreover, since a delayed attack ($t_1 \in (\tau, T]$) is costly and strictly so when $\theta \in \Theta^L$ (assumption 1[A]), and the agent assigns a strictly positive probability that $\theta \in \Theta^L$ (assumption 2[A]), for all $s_i$ and for all $t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}$, we have $(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, t_1)$ for all $t_1 \in (\tau, T]$.

Finally, for any agent with any signal $s_i$, waiting for the disaster alert and then attacking at date $t_1 = \tau$ following $d^\tau = 1$ and attacking at some date $t_0 \in [\tau, T]$ following $d^\tau = 0$ is strictly dominated by attacking immediately at $t_\varnothing = 0$ (or $\mathcal{A}$). Formally, for all $s_i$,

$$(0, t'_0, t'_1) \succ (t_\varnothing, t_0, \tau) \text{ for any } t'_0, t'_1 \in [\tau, T] \cup \{\infty\}, \text{ and } t_\varnothing \in [\tau, T] \cup \{\infty\}, t_0 \in [\tau, T].$$

Consider a plan that involves waiting for the alert (i.e., $t_\varnothing \geq \tau$) and then attacking at time $t_1 = \tau$ following $d^\tau = 1$ and attacking at time $t_0 \in [\tau, T]$ following $d^\tau = 0$. This plan implies the agent will end up attacking regardless of whether $d^\tau = 0$ or 1 and will do so at a date $t \geq \tau$. Compared with the plan of attacking immediately without waiting (i.e., $t_\varnothing = 0$), the delayed attack at any time $t \geq \tau$ yields a lower payoff according to assumption 1(A). It is strictly worse since regardless of the signal $s_i$, the agent believes that $\theta \in \Theta^L$ with a strictly positive probability (assumption 2[A]), and in this event delayed attack is strictly costly (assumption 1[A]). QED

Intuitively, waiting and then attacking after observing the disaster alert $d^\tau$ regardless of whether $d^\tau = 0$ or 1 means that the disaster alert $d^\tau$ is not "informative." Therefore, waiting holds no option value. Since a delayed attack is costly, one should not wait for it. It is important to note that the

above analyses only involve one round of elimination of strictly dominated strategies. This means that as long as the agents are rational, under a disaster alert, regardless of his type $s_i$, an agent either attacks immediately at time 0 or waits for the disaster alert, and if he waits, he follows the disaster alert (attack if and only if the alert is triggered).[18]

Step 2: Perfect Coordination after Disclosure

LEMMA 2. Given that all agents play $\chi_i(s_i) \in \{\mathcal{A}, \mathcal{W}\}$, for any $s_i \in \mathbb{S}$,

(1) if $d^\tau = 1$, all the agents (if any) who have not yet attacked attack immediately at $\tau$, and the game ends in a disaster;
(2) if $d^\tau = 0$, all the agents (if any) who have not yet attacked do not attack afterward, and the game does not end in a disaster.

*Proof.* Recall that if $d^\tau = 1$, then $(\theta, \hat{\mathbf{N}}) \in \mathcal{D}$ for all $\hat{\mathbf{N}} \in \mathbb{N}|\mathbf{N}_\tau$. Therefore, the game ends in a disaster following $d^\tau = 1$. It follows from lemma 1 that if an agent $i$ with signal $s_i$ waits for the disaster alert (i.e., plays $\mathcal{W}$), then conditional on $d^\tau = 1$, the agent attacks at $t_i = \tau$. On the other hand, if $d^\tau = 0$, then $(\theta, \mathbf{N}^0|\mathbf{N}_\tau) \notin \mathcal{D}$. The game may end in a disaster if some agents who waited decide to attack after $d^\tau = 0$. However, it follows from lemma 1 that if an agent $i$ with any signal $s_i$ waits for the disaster alert (i.e., plays $\mathcal{W}$), then conditional on $d^\tau = 0$, the agent never attacks ($t_i = \infty$). Thus, the continuation time path of aggregate attack is $\mathbf{N}^0|\mathbf{N}_\tau$. Hence, the game does not end in a disaster. QED

Lemma 2 demonstrates the effectiveness of the public signal $d^\tau$ in coordinating (rational) agents' actions after the disclosure time $\tau$. Recall that, by design of the disaster alert, no alert means there will be no disaster if no one attacks after the disclosure date $\tau > 0$. Interestingly, lemma 2 shows that, indeed, any agent who chooses to wait for the disaster alert does not attack after seeing no alert, thereby avoiding the disaster. However, it is still possible that agents may not wait for the disaster alert by attacking at time 0.

Step 3: Timely Alert

Lemmas 1 and 2 hold true regardless of the time of disclosure $\tau$ as long as $\tau \in (0, T)$. Next, we explore the timing of disclosure and discuss the role of a "timely" disclosure.

[18] A similar option value argument also appears in other contexts such as in social learning (see, e.g., Chamley and Gale 1994). In their setup, an agent can learn from others' actions, but such actions do not affect his payoff. Consider two agents deciding whether to attack at time 1 or time 2. If an agent waits at a cost to see whether the other agent attacks, it must hold that he will take different actions, conditional on whether the other agent attacks at time 1. Otherwise, waiting has no positive option value.

LEMMA 3. Under assumptions 1 and 2, for any $\tau \in (0, \hat{\tau})$ where $\hat{\tau} = \min\{c^{-1}(g \times \epsilon/(1-\epsilon)), T\} > 0$, under the disclosure policy $\Gamma^\tau$, the only strategy that survives two rounds of iterated elimination of strictly dominated strategies is such that for any $s_i \in \mathbb{S}$, $\chi_i(s_i) = \mathcal{W}$.

*Proof.* We show that under a disaster alert $\Gamma^\tau$ where $\tau$ is sufficiently small, for any signal $s_i$, and any undominated strategy of others $\chi_{-i}$, the expected benefit of playing wait and follow ($\mathcal{W}$) over attacking right away ($\mathcal{A}$) always outweighs the cost.

*Benefit of waiting.*—First, consider the case when the disaster alert is not triggered ($d^\tau = 0$). If the agent plays $\mathcal{W}$, then the realized time of the attack is $\mathrm{t}_i = t_0 = \infty$ (i.e., never attack), whereas under $\mathcal{A}$, the realized time of the attack is $\mathrm{t}_i = t_\varnothing = 0$. Therefore, conditional on $d^\tau = 0$, the conditional expected benefit from playing $\mathcal{W}$, as compared with $\mathcal{A}$, is

$$\mathbb{B}(\Gamma^\tau, \chi_{-i}, s_i) = \mathbb{E}[u(\infty, \theta, \mathbf{N}(\chi_{-i})) - u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 0].$$

Regardless of the undominated strategies others play and agent $i$'s signal $s_i$, the game will not end in a disaster (i.e., $(\theta, \mathbf{N}) \notin \mathcal{D}$) following $d^\tau = 0$ (lemma 2). Therefore, the conditional expected benefit is

$$\begin{aligned} &\mathbb{B}(\Gamma^\tau, \chi_{-i}, s_i) \\ &= \mathbb{E}[u(\infty, \theta, \mathbf{N}(\chi_{-i})) - u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, (\theta, \mathbf{N}(\chi_{-i})) \notin \mathcal{D}] \geq g. \end{aligned} \tag{5}$$

This benefit is from avoiding the mistake of attacking (at time 0) when the disaster does not occur (see [3]), and therefore, it is independent of the disclosure time $\tau$. Moreover, under assumption 2(B), regardless of the signal $s_i$ and regardless of the undominated strategies others play $\chi_{-i}$, the probability of having $d^\tau = 0$ is strictly above $\epsilon$; that is,

$$\mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 0|s_i) \geq \mathbb{P}(\theta \in \Theta^U|s_i) > \epsilon. \tag{6}$$

*Cost of waiting.*—Next, we consider the cost of waiting. Conditional on $d^\tau = 1$, if the agent plays $\mathcal{W}$, then the realized time of attack is $\mathrm{t}_i = t_1 = \tau$, whereas under $\mathcal{A}$, the realized time of attack is $\mathrm{t}_i = t_\varnothing = 0$. Therefore, the conditional expected cost of waiting is

$$\mathbb{C}(\Gamma^\tau, \chi_{-i}, s_i) = \mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i})) - u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 1].$$

By design of the disaster alert $d^\tau$, regardless of the strategy $\chi_{-i}$ others play, the game ends in a disaster (i.e., $(\theta, \mathbf{N}(\chi_{-i})) \in \mathcal{D}$) following $d^\tau = 1$ (lemma 2). Therefore, regardless of $s_i$ and strategy $\chi_{-i}$ others play,

$$\begin{aligned} &\mathbb{C}(\Gamma^\tau, \chi_{-i}, s_i) \\ &= \mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i})) - u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, (\theta, \mathbf{N}(\chi_{-i})) \in \mathcal{D}] \leq c(\tau). \end{aligned} \tag{7}$$

The last inequality comes from assumption 1(B). According to that assumption, the upper bound of the delay cost, $c(\tau)$, continuously increases in the disclosure time $\tau$ and converges to 0 as $\tau \to 0$. The probability of having $d^\tau = 1$ is complementary to the event of $d^\tau = 0$. Under assumption 2(B), it is strictly less than $1 - \epsilon$, regardless of the signal $s_i$ and regardless of the undominated strategies others play $\chi_{-i}$; that is,

$$\begin{aligned}\mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 1|s_i) &= 1 - \mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 0|s_i) \\ &\leq 1 - \mathbb{P}(\theta \in \Theta^U|s_i) < 1 - \epsilon.\end{aligned} \tag{8}$$

*Dominance of $\mathcal{W}$ over $\mathcal{A}$.*—Under any policy $\Gamma^\tau$, the net expected benefit from playing $\mathcal{W}$ as compared with $\mathcal{A}$ for any given $s_i$ and any undominated $\chi_{-i}$ is

$$\begin{aligned}\mathbb{D}(\Gamma^\tau, \chi_{-i}, s_i) \coloneqq\ & \mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 0|s_i)\mathbb{B}(\Gamma^\tau, \chi_{-i}, s_i) \\ & - \mathbb{P}(d^\tau(\theta, \mathbf{N}_\tau(\chi_{-i})) = 1|s_i)\mathbb{C}(\Gamma^\tau, \chi_{-i}, s_i).\end{aligned}$$

Combining inequalities (5)–(8), the following condition holds for any signal $s_i$ and any undominated strategy $\chi_{-i}$ others play:

$$\mathbb{D}(\Gamma^\tau, \chi_{-i}, s_i) > \epsilon \times g - (1 - \epsilon) \times c(\tau).$$

Since $c(\cdot)$ is continuous and strictly increasing, and $c(0) = 0$, there exists

$$\hat{\tau} \coloneqq \min\left\{c^{-1}\left(\frac{\epsilon}{1 - \epsilon} \times g\right), T\right\} \tag{9}$$

such that the (expected) net benefit of waiting $\mathbb{D}(\Gamma^\tau, \chi_{-i}, s_i)$ is strictly positive if the time of disclosure is set at any $\tau \in (0, \hat{\tau})$. Therefore, for any $\tau \in (0, \hat{\tau})$, under policy $\Gamma^\tau$, for any type $s_i$, the choice of $\mathcal{A}$ yields a strictly lower expected payoff than $\mathcal{W}$, irrespective of other agents' strategies (as long as they play undominated strategies). In other words, the only strategy that survives iterated elimination of strictly dominated strategies is to play $\mathcal{W}$ for all $s_i \in \mathbb{S}$. QED

Theorem 1 follows immediately from the above lemmas. If the disaster alert $\Gamma^\tau$ is set in a timely manner (i.e., $\tau \in (0, \hat{\tau})$), under the unique rationalizable strategy profile, regardless of their types $s_i \in \mathbb{S}$, all agents play $\mathcal{W}$ (lemma 3); that is, they wait for the disclosure and then follow the alert afterward. Accordingly, for any fundamental that warrants a run (i.e., $\theta \in \Theta^L$), on path, there is no attack until $\tau$ ($\mathbf{N}_\tau$ is such that $N(t) = 0$ for all $t \in [0, \tau)$), but at $\tau$ the alert gets triggered (since a disaster is inevitable for any $\theta \in \Theta^L$), and all the agents attack at time $\tau$. The game always ends in a disaster. On the other hand, for any fundamental $\theta$ that does not warrant a disaster (i.e., $\theta \notin \Theta^L$), on path, since all the agents wait for the alert, the

alert will not be triggered (i.e., $(\theta, \mathbf{N}^0|\mathbf{N}_\tau) \notin \mathcal{D}$).[19] Then, following $d^\tau = 0$, no agent attacks, and the game does not end in a disaster. Therefore, under the unique rationalizable strategy profile, disaster cannot occur for any $\theta \notin \Theta^L$. In other words, any disaster alert policy $\Gamma^\tau$ with $\tau \in (0, \hat{\tau})$ eliminates panic in the robust sense; that is, $\Theta^P(\Gamma^\tau) = \varnothing$.

Under the timely disaster alert, the agents ignore their private signals and perfectly coordinate their actions: when $\theta \in \Theta^L$, they all attack at time $\tau$; and when $\theta \notin \Theta^L$, they never attack. It is important to note that since all agents wait for the alert, on path, they only learn $\theta \notin \Theta^L$ from the message of "no alert" (or $d^\tau = 0$).

Recall that providing this information at time 0 does not eliminate panic (see sec. III.A). Now, consider an alternative policy, named $\tilde{\Gamma}^\tau$, which discloses, at time $\tau \in (0, T)$, $\tilde{d}^\tau = 1$ if $\theta \in \Theta^L$ and $\tilde{d}^\tau = 0$ otherwise. Under this policy, the agents who wait for the alert will attack if and only if the alert is triggered (i.e., $\tilde{d}^\tau = 1$). In other words, lemma 1 still holds: the strategies that are not strictly eliminated are to take either $\mathcal{A}$ or $\mathcal{W}$ for any signal $s_i$. The critical difference between our proposed $\Gamma^\tau$ and $\tilde{\Gamma}^\tau$ is that if $\theta \notin \Theta^L$ and some agents attack immediately (or they play $\mathcal{A}$), rendering a disaster inevitable, the optimal policy $\Gamma^\tau$ will trigger the alert at time $\tau$, but $\tilde{\Gamma}^\tau$ will not. Therefore, the disaster alert $\tilde{d}^\tau$ under the alternative policy $\tilde{\Gamma}^\tau$ may not predict the occurrence of a disaster (lemma 2 fails). Thus, under $\tilde{\Gamma}^\tau$, conditional on no alert (or $\tilde{d}^\tau = 0$), wait and follow (or playing $\mathcal{W}$) may result in a lower payoff than attacking immediately (or playing $\mathcal{A}$), since the former choice can lead to an inferior outcome of not attacking while the disaster occurs. As the benefit of waiting for the disclosure is not strictly positive, an agent may not wait (lemma 3 fails). Therefore, the choice of attacking immediately (or $\mathcal{A}$) may not eliminated by iterated elimination of strictly dominated strategies for all possible $s_i$. This explains why the principal must combine the information about endogenous attacks with the information about the exogenous fundamentals.

For a timely disaster alert to work, it is crucial that the cost of waiting for the timely alert is small. This is driven by assumption 1(B). This assumption could be violated if, for instance, a disaster occurs immediately following an attack by some agents. Then, even a brief delay could be very costly because it would mean missing the chance to act before the disaster strikes. Accordingly, the agents might not always be willing to wait for the disaster alert, no matter how timely it is. However, assumption 1(B) holds in many well-known applications. In the next section, we provide three such applications.

[19] Otherwise, if $(\theta, \mathbf{N}^0|\mathbf{N}_\tau) \in \mathcal{D}$, then $(\theta, \mathbf{N}) \in \mathcal{D}$ for all possible $\mathbf{N}$ (because $\mathbf{N}^0|\mathbf{N}_\tau$ is such that $N(t) = 0$ for all $t \in [0, T]$, and hence $\mathbf{N} \geq \mathbf{N}^0|\mathbf{N}_\tau$ for all $\mathbf{N}$). By the definition of $\Theta^L$, that implies $\theta \in \Theta^L$, and therefore, this is a contradiction.

## IV. Applications

The timely disaster alert policy closely resembles the early warning system, developed by the World Bank and the IMF, which is designed to assess the fiscal capacity of sovereign debt issuers and to provide advance warnings of potential sovereign defaults. Additionally, bank supervisors implement forward-looking stress tests to evaluate the resilience of banks to future adverse shocks. These policies use a binary disclosure rule ("warning" or "no warning," "pass" or "fail"), with policymakers keen to leverage such disclosures to better coordinate agents' actions so as to avert unwarranted sovereign debt or banking crises.

In this section, we consider three applications of our game—namely, sovereign debt, bank run, and capital outflow—and show that early warnings can be designed in such a way that they avert panic in the robust sense. These applications are well-known in the literature. However, they are usually modeled as static regime change games, whereas we introduce a time dimension to these applications. Complementarity is a common feature across all these applications. Below, we show that in all these applications, the cost of delay satisfies our assumption 1 (B), although for different reasons.

### *A. Sovereign Debt Default*

A country has a significant amount of outstanding sovereign debt that matures in, say, 6 months ($T$). The country experienced an economic recession, and it may default when the sovereign debt matures. This country's fiscal capacity is denoted as $\theta$. Foreign investors who are wary about the recession and resulting currency depreciation may relocate their investments (i.e., attack) before the maturity date. These decisions further worsen the fiscal condition of this country, which contributes to its inability to repay its debt when the debt matures, resulting in a sovereign debt default (or disaster). It is important to note that, in this application, the disaster, if it occurs, can only occur at time $T$.

Suppose that each foreign investor $i \in [0, 1]$ has \$1 invested, and he can relocate it elsewhere at any time, $\mathrm{t}_i \in [0, T]$, or not at all, $\mathrm{t}_i = \infty$. The country defaults on the sovereign debt (or the disaster occurs)—that is, $(\theta, \mathbf{N}) \in \mathcal{D}$—if and only if the aggregate relocation of capital before the maturity date $T$ is at least $\theta$, that is, $N(T) \geq \theta$. The payoff is as follows:

$$u(\mathrm{t}_i, \theta, \mathbf{N}) = \begin{cases} e^{-rt} \times 1 & \text{if} \quad \mathrm{t}_i \in [0, T], \\ 2 & \text{if} \quad \mathrm{t}_i = \infty \text{ and } \theta > N(T), \\ 0 & \text{if} \quad \mathrm{t}_i = \infty \text{ and } \theta \leq N(T). \end{cases}$$

In words, if an investor keeps the investment until the end, then he gets 2 if the country does not default and 0 otherwise. If he relocates his investment

elsewhere, then he is better off by doing so earlier rather than later. An investor loses some benefit from relocation if it is delayed, and we capture this through stationary discounting $r$.

Our disaster alert is an EWS that sends a warning ($d^\tau = 1$) at a specified date $\tau$ if the country will default regardless of what the investors do after time $\tau$ and no warning otherwise.

COROLLARY 1. In the sovereign debt default game, a disaster alert $\Gamma^\tau$, for any $\tau \in (0, \hat{\tau}_1)$ where

$$\hat{\tau}_1 = \min\left\{-\frac{1}{r}\ln\left(1 - \frac{\epsilon}{1-\epsilon}\right), T\right\},$$

eliminates panic (i.e., avoids an unwarranted sovereign debt crisis) in the robust sense.

Note that the cost of delayed attack is $c_1(t) = 1 - e^{-rt}$, which is continuous and strictly increasing, and $c_1(0) = 0$ (assumption 1 holds). The result then immediately follows from theorem 1.

### B. *Dynamic Bank Run*

A bank faces an adverse shock, exemplified by a situation where borrowers within a specific industry begin defaulting on their loans. The bank is safe until the arrival of the shock, but it becomes vulnerable once the shock arrives. Creditors, aware of the imminent arrival of this shock, choose when to withdraw their funds ("run" on the bank, or attack). Once the shock hits, the bank can fail if it is not strong enough to withstand the aggregate withdrawal. Unlike in the first example, the bank may fail at any date in $[0, T]$ and not just at time $T$.[20] A creditor prefers attacking before the bank fails rather than waiting until after it fails. This first-mover advantage may make delayed attacks significantly costly. However, if the distribution of the shock arrival time has no jumps, the cost has a continuous upper bound; that is, assumption 1 holds. Therefore, a timely disaster alert will eliminate panic-based bank runs. Below, we formalize this argument.

For any $(\theta, \mathbf{N})$, an agent's payoff from $\mathrm{t}_i \in [0, T] \cup \{\infty\}$ is stochastic since it also depends on when the shock arrives. The arrival time of the adverse shock, $t_s$, follows a distribution $G$ on $[0, T]$, where $G(t)$ is continuously (strictly) increasing in $t$, $G(0) = 0$, and $G(T) = 1$. The stochastic payoff is denoted by $\pi(\mathrm{t}_i; \theta, \mathbf{N}, t_s)$, and the agent's expected payoff is

$$u(\mathrm{t}_i, \theta, \mathbf{N}) = \mathbb{E}_G[\pi(\mathrm{t}_i; \theta, \mathbf{N}, t_s)].$$

[20] Although for this application, we assume that the bank can fail instantaneously, in practice, bank payments and fund redemption are often systematically scheduled to clear at a fixed time, such as 4 p.m. each business day. Therefore, if we interpret the time unit in our model as minutes or hours and assume that disaster alerts can be processed and disclosed very rapidly, then the feature that disasters can occur only at a fixed "date," as in sec. IV.A, is also valid for bank runs. Therefore, the same result applies to such bank runs.

Only after the shock arrives does the bank become vulnerable and, thus, subject to failure. Before the adverse shock arrives ($t < t_s$), the bank's fundamental is $\theta_0 \in \Theta^U$, and therefore, the bank is not vulnerable to runs, whereas the postshock fundamental is $\theta$. We define the strength of the bank's balance sheet as $R(\theta, N)$, which is a continuously differentiable function. A better fundamental $\theta$ makes the bank stronger (i.e., $(\partial R/\partial \theta) > 0$) whereas more attacks weaken the bank's strength (i.e., $(\partial R/\partial N) < 0$). The bank fails if $R(\theta, N(T)) \leq 0$, and it fails at the earliest point of time $t$ when two conditions hold: (1) the adverse shock has arrived (i.e., $t \geq t_s$) and (2) the bank's strength is exhausted (i.e., $R(\theta, N(t)) \leq 0$). Formally, we define[21]

$$t_R(\theta, \mathbf{N}) := \begin{cases} \min\{t | R(\theta, N(t)) \leq 0\} & \text{if } R(\theta, N(T)) \leq 0, \\ \infty & \text{if } R(\theta, N(T)) > 0; \end{cases}$$

and define the time when the bank fails as

$$t_f(\theta, \mathbf{N}, t_s) := \max\{t_s, t_R(\theta, \mathbf{N})\}.$$

If $t_s > t_R$, it becomes evident at time $t = t_R$ that the bank will fail as soon as the shock arrives.[22]

If an agent chooses to attack and does so at time $t_i = t \in [0, T]$, then he gets a payoff of

$$\pi(t; \theta, \mathbf{N}, t_s) = \mathbb{1}(t \leq t_f(\theta, \mathbf{N}, t_s))\bar{U} + \mathbb{1}(t > t_f(\theta, \mathbf{N}, t_s))\underline{U}.$$

The above specification captures the first-mover advantage. The payoff from attacking depends on the agent's position in the queue of attacking agents. If he attacks early enough, before the time of bank failure ($t_i \leq t_f$), he obtains a high payoff $\bar{U}$; otherwise, he is too late in attacking ($t_i > t_f$) and therefore obtains a low payoff $\underline{U}$. On the other hand, if an agent does not attack at all ($t_i = \infty$), then he gets a payoff of

$$\pi(\infty; \theta, \mathbf{N}, t_s) = \mathbb{1}(t_f(\theta, \mathbf{N}, t_s) = \infty)\bar{V} + \mathbb{1}(t_f(\theta, \mathbf{N}, t_s) < \infty)\underline{V}.$$

That is, if the bank does not fail ($t_f = \infty$), he gets a high payoff $\bar{V}$, and if the bank fails ($t_f < \infty$), he gets a lower payoff $\underline{V}$. We assume $\bar{V} > \bar{U} > \underline{U} \geq \underline{V}$.

Note that given the stochastic nature of the shock's arrival, the ability to schedule disclosures ahead of the shock's actual occurrence makes the policy

[21] Note that, by definition, if the bank does not fail by the end, then $t_f = \infty$; and if the bank fails (i.e., $R(\theta, N(T)) \leq 0$), since both functions $R$ and $N$ are measurable ones, then $[t_s, T] \cap \{t | R(\theta, N(t)) \leq 0\}$ is a closed interval, and therefore the minimum of this interval (or $t_f$) is well defined.

[22] As in our benchmark setup, we maintain the assumption that the agents do not see any new information over time. However, a natural assumption in this application is that the agents will see when the bank fails. If we allow the agents to observe when the disaster occurs, we need to modify the strategy space that allows the agents to attack whenever they see that the disaster occurs. This only makes our result stronger (see sec. V.C for details).

forward looking.[23] Our disaster alert is a forward-looking stress test, which, at a specified date $\tau$, gives the bank a "fail" grade ($d^\tau = 1$) if it is inevitable that the bank will fail (if it has not already) regardless of what the creditors do starting from time $\tau$ and a "pass" grade ($d^\tau = 0$) otherwise. Passing the stress test means that the bank can sustain the forthcoming adverse shock in the absence of any further (endogenous) attacks from the creditors.

COROLLARY 2. In the dynamic bank run game, with the stochastic arrival of shock $G(t)$, where $G(t)$ is continuously (strictly) increasing in $t$, $G(0) = 0$ and $G(T) = 1$, a forward-looking bank stress test $\Gamma^\tau$, with $\tau \in (0, \hat{\tau}_2)$ where

$$\hat{\tau}_2 = G^{-1}\left(\min\left\{\frac{\epsilon}{1-\epsilon} \times \frac{\bar{V} - \bar{U}}{\bar{U} - \underline{U}}, 1\right\}\right) \tag{10}$$

eliminates panic (i.e., prevents unwarranted bank runs) in the robust sense.

Note that a delayed attack is costly since the agent may get $\underline{U}$ instead of $\bar{U}$ if the bank fails in the meantime. If $\theta \in \Theta^L$, then the bank fails as soon as the shock arrives, which makes a delayed attack strictly costly (since $G$ is strictly increasing). The expected cost of a delayed attack is bounded above by $c_2(t) = G(t)(\bar{U} - \underline{U})$. Since $G(t)$ is continuous and $G(0) = 0$, assumption 1(B) holds. The result then immediately follows from theorem 1.

It is worth noting that the stochastic arrival of adverse shocks is necessary for the bank run application, wherein a bank can fail instantaneously following a significant wave of runs at any given time, and running before bank failure is advantageous over running after (i.e., the first-mover advantage). To illustrate this, consider a scenario where the shock is certain to arrive before the game begins. In such a case, if agents attack immediately, then the bank is very likely to fail at time 0. As such, waiting for the disaster alert will cost $(\bar{U} - \underline{U})$ regardless of $\tau$, and accordingly, an agent may not wait for the disaster alert. Under the stochastic arrival of adverse shocks, the (expected) waiting cost can be small under a timely disclosure policy since the likelihood that the waiting results in the loss of an opportunity to run before the bank failure is continuous in the duration of waiting.[24]

[23] Forward-looking stress tests require due diligence. In sec. II.1 of the online appendix, we discuss how the first bank stress test was conducted in the United States in 2009 and draw parallels with our proposed policy. In sec. II.2 of the online appendix, we discuss the case in which the bank regulator (principal) may not be able to forecast the fundamental $\theta$ accurately. If the principal makes a mistake and triggers an alert even though a disaster is not inevitable (false negative), then all the agents will attack, possibly causing a disaster. On the other hand, if the principal makes false positive mistakes, then agents may not find the alert very valuable and may decide not to wait for it. The false positive may also arise not because of the principal's mistake but because some agents do not behave rationally. In proposition II.1, we show that if the false positive probability is not sufficiently large, our main result holds.

[24] Suppose that an agent gets $\underline{U}$ from attacking exactly when the bank fails ($t = t_f$); i.e., the first-mover advantage exists for times that are strictly ahead of (instead of weakly ahead of) the regime change time. Then, even by attacking at time 0, an agent will sometimes miss the opportunity to get the high payoff ($\bar{U}$). Conditional on the shock arrival time being $t_s = 0$, if $d^\tau = 1$,

Our analysis highlights the effectiveness of forward-looking bank stress tests, as a disclosure policy, in preventing unnecessary bank panics.[25] If the hypothetical adverse shock considered in the stress test accurately reflects the impending fundamental shock (or $\theta$), then the test results—pass ($d = 0$) or fail ($d = 1$)—scheduled for disclosure at time $\tau$ (even before the actual onset of the adverse shock) can persuade bank creditors to disregard their private information, wait for the stress test result, and then coordinate on following it.

Furthermore, our analysis emphasizes that the timeliness of disclosing the stress test results is critical for its efficacy.[26] Our model also draws empirical implications for the relationship between the timeliness and efficacy of stress tests. From the determination of the disclosure time's upper limit $\hat{\tau}_2$ (see [10]), for a stress test to be effective, the disclosure time should be set earlier if the market confidence in the bank's fundamentals is lower (i.e., the lower bound of $\mathbb{P}(\theta \in \Theta^U | s_i)$, or $\epsilon$, is smaller), the reward of staying $\bar{U}$ (e.g., the interest payment) is higher, and/or the bankruptcy payoff $\underline{U}$ (e.g., deposit insurance coverage) is lower.

### C. *Slow-Moving Capital Outflow*

A mass of investors decide when to attack (relocate their capital), and a disaster occurs (the market collapses) when the aggregate attack becomes so large that the fundamental $\theta$ is not good enough to sustain it. In practice, moving capital away from a country or withdrawing funds from financial institutions often takes time. This could be due to some exogenous constraints. For example, there are constraints on cross-board capital flows. Withdrawing from money market mutual funds or hedge funds might be subject to redemption gates. Or it can happen simply because moving capital away, which involves selling off properties and firing employees, is time consuming. Unlike in the first application, the disaster

---

then the regime changes at time 0. In that case, there is no difference between the payoff from $\mathcal{A}$ and $\mathcal{W}$. This reduces the cost of waiting to $\tilde{c}_2(t) = (G(t) - G(0))(\bar{U} - \underline{U})$. It then follows from the continuity of $G(\cdot)$ that $\tilde{c}_2(t) \to 0$ as $t \to 0$ regardless of $G(0)$. Thus, it is easier to persuade the agents to wait in this alternative setup, where the assumption $G(0) = 0$ can be relaxed.

[25] The bank stress test serves many other purposes and is often designed in conjunction with other supplementary policies. See Goldstein and Leitner (2018) for how such disclosure policies can make it easier for banks of heterogeneous quality to raise funds under adverse selection. See Orlov, Zryumov, and Skrzypacz (2023) for how it helps to uncover systemic risk and how it can be combined with capital regulations. See Inostroza (2019) for a comprehensive disclosure policy including banks' liquidity position and asset quality. In this paper, however, we focus only on the information design aspect of the policy.

[26] Anderson (2016) presents anecdotal evidence to support the importance of conducting bank stress tests in a timely manner. The author illustrates that forward-looking stress tests proved to be highly effective in the United States but less so in Europe. This disparity, according to the author, can be attributed to the timeliness of US stress tests compared with their European counterparts, which were arguably conducted "too late."

can happen at any time, and unlike in the second application, the shock has no stochastic arrival time. Nevertheless, the cost of a delayed attack can have a continuous upper bound if the market moves slowly. Therefore, a timely disaster alert eliminates investor panic in such markets. Below, we formalize this argument.

A mass of investors decide when to attack, and it takes $l > 0$ units of time to collect the entire payoff from an attack. We assume that any agent who has initiated an attack at $t_i$ finishes the $B((t - t_i)/l)$ part of his attack by time $t$, where $B(\cdot)$ has support $[0, 1]$ and admits a density $b(\cdot)$, which is Lipschitz continuous; that is, for any $t' \neq t$, $|b(t) - b(t')| \leq \kappa |t - t'|$ for some positive constant $\kappa$. The cumulative pledged attack until time $t$ is $N(t)$, and the materialized attack until time $t$ is $M(t)$, where $M(t) := \int_{\nu \in [0,t]} B((t - \nu)/l) \mathrm{d}N(\nu)$. By definition, $M(t) \leq N(t)$ and $M(T + l) = N(T)$.

The game ends in a disaster (i.e., $(\theta, \mathbf{N}) \in \mathcal{D}$) if, and only if, $R(\theta, N(T)) = R(\theta, M(T + l)) \leq 0$, where $R(\cdot)$ has the same properties as in section IV.B. By any time $t$, whether a disaster has occurred or not depends on the materialized attack $M(t)$ instead of the pledged attack $N(t)$, that is, whether or not $R(\theta, M(t)) \leq 0$. We define $t_f^M$ as the first instance when disaster happens; that is,

$$t_f^M(\theta, \mathbf{N}) := \begin{cases} \min\{t | R(\theta, M(t)) \leq 0\} & \text{if} \quad R(\theta, N(T)) \leq 0; \\ \infty & \text{if} \quad R(\theta, N(T)) > 0. \end{cases}$$

An agent who initiates an attack at time $t_i$ receives a high *flow payoff* $\bar{U}$ at any instant $t \in [t_i, t_i + l]$ before the disaster occurs ($t \leq t_f^M$) and a low flow payoff $\underline{U} \in (0, \bar{U})$ after the disaster occurs. Moreover, similar to the first example, we assume that an investor loses some benefit from leaving the market late and investing elsewhere, and we capture this assumption through stationary discounting. Therefore, the payoff from starting an attack at time $t \in [0, T]$ is

$$u(t, \theta, \mathbf{N}) = \int_t^{t+l} e^{-rz} \left[ \mathbb{1}(z \leq t_f^M(\theta, \mathbf{N})) \bar{U} + \mathbb{1}(z > t_f^M(\theta, \mathbf{N})) \underline{U} \right] \mathrm{d}B\left(\frac{z - t}{l}\right).$$

As in section IV.B, we assume that if an agent does not attack at all ($t_i = \infty$), then he gets a payoff of

$$u(\infty, \theta, \mathbf{N}) = \mathbb{1}(t_f^M(\theta, \mathbf{N}) = \infty) \bar{V} + \mathbb{1}(t_f^M(\theta, \mathbf{N}) < \infty) \underline{V}.$$

Finally, we assume $\bar{V} > \bar{U}$ and $e^{-r(T+l)} \underline{U} > \underline{V} \geq 0$.

The disaster alert policy, $\Gamma^\tau$, at time $\tau \in (0, T)$, triggers the alert if at time $\tau$ it becomes evident that the disaster is inevitable, regardless of what the agents (who have not yet initiated an attack) do starting from time $\tau$, and does not trigger the alert otherwise.

COROLLARY 3. Suppose that capital is slow moving; that is, if an agent initiates an attack at time $t_i$, he finishes the $B((t - t_i)/l)$ fraction of

attack by time $t$, where $l > 0$ and the associated density $b(\cdot)$ is Lipschitz continuous (i.e., $|b(t) - b(t')| \leq \kappa |t - t'|$ for a positive constant $\kappa$). Then, under policy $\Gamma^{\tau}$ with $\tau \in (0, \hat{\tau}_3)$, where

$$\hat{\tau}_3 = \min\left\{c_3^{-1}\left(\frac{\epsilon}{1-\epsilon}(\bar{V} - \bar{U})\right), l, T\right\}, \quad \text{and} \quad c_3(t) = \bar{U}\left(B\left(\frac{t}{l}\right) + \frac{\kappa t}{l}\right),$$

panic is eliminated in the robust sense.

In this application, a delayed attack is strictly costly since $\bar{U} > \underline{U} > 0$ and $r > 0$. In the proof of corollary 3, we show that $c_3(t)$ works as a continuous upper bound on the cost of a delayed attack. To see this, consider a small delay $t < l$. The difference in the part of the attack that is executed between an immediate attack at time 0 and a delayed attack at time $t \in (0, l)$ is (1) $B(t/l)$ versus 0 in the time interval $[0, t]$; (2) $\int_{z=t}^{l} dB(z/l)$ versus $\int_{z=t}^{l} dB((z-t)/l)$ in the time interval $[t, l]$; and (3) 0 versus $\int_{z=l}^{t+l} dB((z-t)/l)$ in the time interval $(l, t + l]$, which is nonpositive. Since $b(\cdot)$ is Lipschitz continuous, the second difference in the part of the attack that can be finished in $[t, l]$ is at most $\kappa t/l$. Recall that the most benefit an agent can get by finishing his attack is $\bar{U}$ (since $\bar{U} > \underline{U} > 0$). This gives us the upper bound on the cost of a delayed attack $c_3(t)$, which is continuously increasing in $t$ with $c_3(0) = 0$. The result then immediately follows from theorem 1.

Recall that only under assumption 1(B)—the (expected) cost of delay can be bounded above by a function that is continuous in the duration of waiting—can the timely disaster alert policy work to eliminate panic in a robust sense. These three examples demonstrate that it is a reasonable condition that fits into the dynamic environment of each application. In the first application, this condition holds naturally because the disaster can only occur at the end (at time $T$). When the disaster can occur before time $T$, attacking earlier than the occurrence of disaster yields a significant first-mover advantage. However, if the disaster cannot occur before the stochastic arrival of a negative shock (as in our second application), then the probability of a disaster occurring while an agent waits for a short duration is limited, making the cost of a delayed attack small. The cost could also be small if, for instance, it takes time for the attack to be fully executed (as in our third application).

## V. Extensions

In this section, we discuss how our main result works with a more flexible information structure. In section V.A, we use the simple two-player example mentioned in the introduction to discuss how the doubt assumption can be relaxed. In section V.B, we consider a global game perturbation (Gaussian noisy signal) to show that the argument can be extended. Sections V.C and V.D extend the model to consider the cases in which the agents can observe the occurrence of a disaster (if it happens) and

can receive additional information about the fundamental $\theta$ over time, respectively.

### A. *Doubt Assumption*

Thus far, we have maintained the assumption that, regardless of their private signals, the agents always have doubt (assumption 2). It is essential for the option value argument (see lemma 1) that regardless of his type, each agent assigns a strictly positive probability that $\theta \in \Theta^L$. This is guaranteed by assumption 2(A); that is, $\mathbb{P}(\theta \in \Theta^L | s_i) > 0$ for all possible $s_i$. On the other hand, to ensure that there is a lower bound on the expected benefit of waiting for the disaster alert (which is independent of what strategy other agents play), we adopt assumption 2(B), under which $\mathbb{P}(\theta \in \Theta^U | s_i) > \epsilon > 0$ for all possible $s_i$.

In this section, we argue that assumption 2(B) is not necessary for our main result. We show that when agents have heterogeneous beliefs, our result may hold even when assumption 2(B) does not hold for some types.

Intuitively, an agent whose belief violates assumption 2(B) lacks a lower bound on the benefits of delaying action. In extreme cases, the agent is certain that the state $\theta$ does not fall within the upper dominance region, making waiting purely costly if all others opt to attack immediately. However, with heterogeneous beliefs, this agent may believe that assumption 2(B) holds for others and, therefore, they will not attack immediately. This higher-order belief can make waiting beneficial even when $\theta$ lies outside the upper dominance region, and thus, the agent might also want to wait. Below, we incorporate heterogeneous beliefs into the two-player, two-period example from the introduction to formalize this idea.

Consider the following information structure. Nature is equally likely to choose the state $\theta \in \{B, M, G\}$, and conditional on the state, nature independently draws a signal $s_i \in \{l, m, h\}$ for each player $i$ from the following conditional distribution with $p \in (0.5, 1)$:

| $\mathbb{P}(s_i \mid \theta)$ | $l$ | $m$ | $h$ |
|---|---|---|---|
| $\theta = B$ | $p$ | $\frac{1}{2}(1-p)$ | $\frac{1}{2}(1-p)$ |
| $\theta = M$ | $\frac{1}{2}(1-p)$ | $p$ | $\frac{1}{2}(1-p)$ |
| $\theta = G$ | 0 | $1-p$ | $p$ |

Note that under this information environment, assumption 2(A) holds since $\mathbb{P}(\theta = B | s_i) > 0$ for $s_i = l, m, h$. However, if an agent receives the signal $l$, then he knows for sure that the fundamental is not in the upper dominance region. Hence, assumption 2(B) is violated for $s_i = l$.

Claim 1. Under the above information structure (which violates assumption 2[B]), if the cost of waiting is sufficiently small—that is, $c < (1-p)/(1+p)$—then under a period-2 disaster alert, the unique

rationalizable strategy is to wait and follow (*WNA*) for all possible $s_i \in \{l, m, h\}$, and therefore, panic is eliminated in the robust sense.

To understand this result, first note that, since $\mathbb{P}(\theta = B|s_i) > 0$ holds for $s_i = l, m, h$, the only strategies that survive the first round of elimination are *A* and *WNA* for all possible $s_i$. Since assumption 2 holds for $s_i = m, h$, we can apply the result we discussed in the introduction to show that, under the condition $c < 2(1 - p)/(1 + p)$, the only strategy that survives the second round of elimination for $s_i = m, h$ is wait and follow (*WNA*).[27] However, since player *i* with $s_i = l$ assigns zero probability to $\theta = G$ (assumption 2[B] fails), we cannot eliminate strategy *A* for $s_i = l$ in the second round.

Note that after two rounds of elimination of strictly dominated strategies, player *i* with $s_i = l$ understands that, under the condition $c < 2(1 - p)/(1 + p)$, if the opponent is rational, then he will take *WNA* when receiving either $s_j = m$ or $s_j = h$. Therefore, when the state is $\theta = M$ and when the opponent receives either $s_j = m$ or *h*, if player *i* does not attack, then the alert will not be triggered ($d = 0$) and the opponent will not attack after seeing $d = 0$. Provided that there is no-alert ($d = 0$), attacking immediately (*A*) yields a payoff of 1, whereas playing *WNA* yields a payoff of 2. Note that, conditional on $s_i = l$ and the belief that the opponent is rational, the event $d = 0$ arises with probability[28]

$$\begin{aligned}\mathbb{P}(d = 0|s_i = l) &\geq \mathbb{P}(s_j \in \{m, h\}|\,\theta = M)\,\mathbb{P}(\theta = M|s_i = l)\\ &= \left(p + \frac{1 - p}{2}\right)\frac{1 - p}{1 + p} = \frac{1 - p}{2} > 0;\end{aligned} \tag{11}$$

and with probability $(1-((1 - p)/2))$, the state is *B* or the state is *M* while the opponent receives $s_j = l$. Therefore, by playing *WNA*, he gets at least (if we assume the opponent plays *A* with probability 1 when $s_j = l$) $((1 - p)/2) \times 2+(1-((1 - p)/2))(1 - c)$, which is strictly better than attacking right away under the condition $c < (1 - p)/(1 + p)$. Therefore, provided that $c < (1 - p)/(1 + p)$, the unique strategy that survives three rounds of an iterated elimination of strictly dominated strategies is to wait and follow (*WNA*) for all possible types $s_i \in \{l, m, h\}$. Therefore, panic is eliminated in the robust sense.

### *B. Global Game Information Structure*

The above simple example illustrates that the doubt assumption can be weakened when the agents have heterogeneous beliefs. The idea is that,

[27] Note that, since $p > 1/2$, $\mathbb{P}(\theta = G|s_i = h) = p > \mathbb{P}(\theta = G|s_i = m) = (2 - 2p)/(3 - p)$. Therefore, to get the sufficient condition for the waiting cost, we can replace $\pi_G$ with $(2 - 2p)/(3 - p)$ in the condition $c < \pi_G/(1 - \pi_G)$.

[28] Note that the event of no alert also arises if the state is $\theta = M$ and the opponent plays *WNA* after receiving signal $s_j = l$. This explains the first inequality in (11).

in the game with strategic complementarity, even when an agent does not have doubt—that is, he does not believe that the fundamental is in the upper dominance region (so that waiting cannot be beneficial when all others attack immediately)—he would still prefer waiting and following the principal's recommendation if he believes that others have doubt (so that they will play wait and follow) or that others believe that others have doubt (so that they will wait and follow because they believe that others will do the same), and so on.

In this section, we consider a global game information structure. For tractability, we resort to a parametric example. We take our bank run example from section IV.B and assume that the regime strength $R(\theta, N(t)) = \theta - N(t)$. We specialize to a Gaussian environment widely used in the global game literature. Nature draws $\theta$ from the improper prior $\mathcal{U}[-\infty, \infty]$, and agents receive conditionally independent noisy signal $s_i = \theta + \sigma \varepsilon_i$, where $\varepsilon_i \sim^{\text{iid}} N(0, 1)$ and $\sigma > 0$ is a scale parameter. In this specialized environment, $\Theta^L = (-\infty, 0]$ and $\Theta^U = (1, \infty)$, and for any signal $s_i$, $\mathbb{P}(\theta \in \Theta^L | s_i) > 0$ and $\mathbb{P}(\theta \in \Theta^U | s_i) > 0$. Given that assumption 2(A) holds, the only strategies that survive the first round of iterated elimination are such that, for any possible $s_i$, agents choose between $\mathcal{A}$ and $\mathcal{W}$ (see the proof of lemma 1). However, note that, for any given $\epsilon$, an agent who receives signal $s_i < 1 + \sigma \Phi^{-1}(\epsilon)$ assigns a probability of less than $\epsilon$ to $\theta \in \Theta^U$, which violates assumption 2(B).

PROPOSITION 1. In the above specialized environment, the following holds:

1. Under no disclosure, the unique rationalizable strategy is to attack at $t = 0$ for $s_i \leq \hat{s} = \hat{p} + \sigma \Phi^{-1}(\hat{p})$ and never attack for $s_i > \hat{s}$, and accordingly, the regime does not change if and only if $\theta > \hat{\theta} = \hat{p}$, in which

$$\hat{p} = \frac{\bar{U} - \underline{V}}{\bar{V} - \underline{V}} \in (0, 1).$$

2. Under disaster alert policy $\Gamma^\tau$ with any $\tau \in (0, T)$, the unique rationalizable strategy is to attack at $t = 0$ (or play $\mathcal{A}$) for $s_i \leq s^* = p^*(\tau) + \sigma \Phi^{-1}(p^*(\tau))$ and to wait and follow (or play $\mathcal{W}$) for $s_i > s^*$, and accordingly, the regime does not change if and only if $\theta > \theta^*(\tau) = p^*(\tau)$, in which

$$p^*(\tau) = \frac{\bar{U} - \left( G(\tau)\underline{U} + (1 - G(\tau))\bar{U} \right)}{\bar{V} - \left( G(\tau)\underline{U} + (1 - G(\tau))\bar{U} \right)} \in (0, \hat{p}).$$

3. Panic occurs under no disclosure when $\theta \in (0, \hat{\theta}]$, and it occurs under the disaster alert when $\theta \in \Theta^{p}(\Gamma^{\tau}) = (0, \theta^{*}(\tau)]$. The policy $\Gamma^{\tau}$ reduces the probability of panic; that is, $\theta^{*}(\tau) < \hat{\theta}$ for any $\tau \in (0, T)$.
4. An earlier disaster alert is more effective; that is, $\theta^{*}(\tau)$ falls as $\tau$ decreases. For any $\zeta > 0$ (however small), there exists

$$\hat{\tau}(\zeta) = G^{-1}\left(\min\left\{\frac{\bar{V} - \bar{U}}{\bar{U} - \underline{U}}\frac{\zeta}{1 - \zeta}, 1\right\}\right) > 0$$

such that, under the policy $\Gamma^{\tau}$ with $\tau \in (0, \hat{\tau}(\zeta))$, $\theta^{*}(\tau) < \zeta$.

Recall that, under no disclosure, the game boils down to a static regime change game, where the agents choose between "attack" and "not attack." Morris and Shin (2003) use the infection argument and show that for each type $s_i$, a unique strategy survives the iterated elimination of dominated strategies: attack for $s_i \leq \hat{s}$ and not attack for $s_i > \hat{s}$. This implies that the regime survives if and only if the fundamental is beyond some cutoff $\hat{\theta}$. This cutoff is $\hat{\theta} = \hat{p}$ (see the appendix for the formal argument). Notice that $\hat{p}$ is the probability of survival that makes an agent indifferent between attacking and not attacking.

Under a disaster alert policy, the difference is that, instead of a "not attack" strategy, the agents play $\mathcal{W}$—that is, they wait and (do not) attack if the disaster alert is (not) triggered (see lemma 1). Clearly, an agent gets a higher payoff from $\mathcal{W}$ than from the choice of "not attack" under no disclosure as the strategy $\mathcal{W}$ provides the agents with an opportunity to attack later when the regime change becomes inevitable. Therefore, when the option of waiting for disclosure is available, agents are less likely to attack the regime at time 0. Accordingly, based on the unique strategy that survives iterated elimination of strictly dominated strategies, the fundamental cutoff $\theta^{*}(\tau)$ is lower than the fundamental cutoff $\hat{\theta}$ in the case without a disclosure policy. This implies that the disaster alert policy helps avert some panic if it is set at any date $\tau \in (0, T)$. As $\tau$ decreases, the expected payoff from $\mathcal{W}$ further increases and, accordingly, $\theta^{*}(\tau)$ falls. Proposition 1 shows that the fundamental cutoff $\theta^{*}(\tau) \to 0$ and the panic set $\Theta^{p}(\Gamma^{\tau})$ converges to an empty set as $\tau$ becomes almost 0.

As in the canonical regime change game, in the above proposition, the global game infection argument selects a unique rationalizable strategy profile, which is independent of the scale parameter of the noise (or $\sigma$). That is, even when the agents are almost certain about the underlying fundamental ($\sigma$ is close to 0), a disaster alert can persuade them to wait and follow (in the limit as $\tau$ becomes almost 0). In contrast, any static disclosure will be completely ineffective when the agents are almost sure about the fundamental, as it only discloses information about $\theta$. On the other hand, the disaster alert perfectly predicts whether or not the regime

will change (lemma 2), regardless of whether the fundamental uncertainty (measured by $\sigma$) is large or small. As long as there is fundamental uncertainty (i.e., $\sigma > 0$), the disaster alert provides valuable information about the regime outcome, thus incentivizing agents to wait and perfectly coordinate their actions after the disclosure.

### C. *Observing When the Disaster Begins*

In sections IV.B and IV.C, we looked into applications where the disaster may begin anytime within the time window $[0, T]$. A natural assumption in these applications is that the agents observe when the disaster begins. However, for expositional simplicity, we maintain the same assumption as in the main section: the agents do not observe any new information over time. Below, we discuss why our result will be even stronger if the agents can observe the occurrence of the disaster (if that happens).

Consider, for instance, the bank run application. When the agents can observe the bank failure, under a disaster alert policy $\Gamma^\tau$, the only rationalizable strategies are (1) attack right away ($\mathcal{A}$) and (2) attack as soon as the regime has changed before time $\tau$, or, if the regime has not changed by time $\tau$, attack at time $\tau$ when the alert is triggered (i.e., $d^\tau = 1$) and do not attack if the alert is not triggered (i.e., $d^\tau = 0$). Let us call this second strategy $\mathcal{W}_o$. The difference between strategies $\mathcal{W}_o$ and $\mathcal{W}$ is that, with $\mathcal{W}_o$, the agent, during the waiting period, can attack at an earlier date if the bank fails before the disclosure time $\tau$. Thus, compared with the case in which the bank failure cannot be observed, the expected cost of delay decreases. As such, it is even easier to dissuade the agents from playing $\mathcal{A}$. Under the same timely disaster alert policy $\Gamma^\tau$ with $\tau \in (0, \hat{\tau})$, all the agents play strategy $\mathcal{W}_o$, and thus panic is eliminated even when the disaster is publicly observable.

### D. *Arrival of Additional Information*

In our benchmark setup, we assume that the principal has full control of the information after the game begins, and the agents do not get any additional information over time. In practice, this may not always be the case, especially if the time horizon is long. For example, if the time horizon is a month, agents may receive weekly updates about the fundamental $\theta$ from other sources.

Suppose that an agent $i$ receives a vector of noisy signals $\mathbf{s}_i := (s_i^0, s_i^1 \ldots, s_i^K)$ at specified dates $\{\nu_0, \nu_1, \ldots \nu_K\}$, where $\nu_0 = 0$ and $\nu_K < T$. As before, the noise can be arbitrarily correlated, and regardless of $\mathbf{s}_i$, an agent $i$ always has doubt.[29] We make two additional assumptions. First,

[29] It is worth noting that a practical scenario where different groups of agents receive noisy information about $\theta$ at different times serves as a specific example of this general setting. In

we generalize the assumption of waiting cost as follows. There exists a continuous and strictly increasing function $c(t)$ defined on $[0, T]$ with $c(0) = 0$ such that, for any $(\theta, \mathbf{N}) \in \mathcal{D}$, $t$, $t' \in [0, T]$ and $t < t'$, $u(t, \theta, \mathbf{N}) - u(t', \theta, \mathbf{N}) \leq c(t') - c(t)$. Second, we assume that the information arrives infrequently; that is,

$$\bar{\tau} := \min_{j=1}^{K+1}\{\nu_j - \nu_{j-1}\} > 0,$$

where $\nu_{K+1} = T$. This implies that whenever the agents receive some information, there is at least a time window of $\bar{\tau}$ before new information arrives or the game ends. The fact that $\bar{\tau} > 0$ gives the principal the scope to set a timely disaster alert shortly after each arrival of new information and before the next information arrives or the game ends. We construct an extended disaster alert policy as follows: a timely disaster alert is made every time after new information arrives. That is, the $(k + 1)$th disaster alert is triggered:

$$d^{\nu_k+\tau}(\theta, \mathbf{N}_{\nu_k+\tau}) = \begin{cases} 1 & \text{if } (\theta, \mathbf{N}^0|\mathbf{N}_{\nu_k+\tau}) \in \mathcal{D} \\ 0 & \text{otherwise} \end{cases}, k = 0, 1, \ldots K.$$

Proposition 2. If exogenous information arrives infrequently over time, there exists $\tilde{\tau} > 0$ such that, under the extended disclosure policy $\Gamma^\tau$ with $\tau \in (0, \min\{\tilde{\tau}, \bar{\tau}\})$, the unique rationalizable strategy is to wait for all the alerts and then follow them for all possible $\mathbf{s}_i$. Under this strategy, panic is eliminated in the robust sense; that is, $\Theta^P(\Gamma^\tau) = \varnothing$.

When the agents get additional information over time, one disaster alert $d^\tau$ is not enough to be persuasive. An agent may act on his new information arriving later than the disaster alert and attack at that time. As proposition 2 states, however, as long as the principal can set a timely disaster alert every time after the arrival of new information, the agents will never act on their own signals. Instead, they will always choose to wait for the next disclosure and follow the disaster alert.[30] We prove this by extending the iterated elimination argument from theorem 1. We inductively show that any strategy involving waiting for the alerts before time $\nu_k$ $(k = 0, 1, \ldots, K)$, then attacking at time $\nu_k$ but not waiting for the later alert(s) is strictly dominated by the strategy of waiting for all alerts and

this case, group $k + 1$ $(k = 0, 1, \ldots, K)$ agents receive informative signals $s_i^k$ only at time $\nu_k$. At all other times, they do not receive informative signals (i.e., signals $s_i^{k'}$ for any $k' \neq k$ are not informative about $\theta$).

[30] This result provides a rationale for the prevalence of periodic disclosures such as DFAST (Dodd-Frank Act Stress Test) and CCAR (Comprehensive Capital Analysis and Review) stress tests. We thank an anonymous referee for pointing this out.

never attacking unless any alert is triggered. The formal proof is relegated to the appendix.

## VI. Discussion and Conclusion

Early warnings, when designed appropriately, can be remarkably effective under a reasonably general setup. Our proposed policy resembles an EWS, such as debt sustainability analysis or a forward-looking bank stress test, which cautions agents about the impending crisis. This paper provides a rationale for adopting such a policy by uncovering the underlying mechanism that makes an EWS effective at preventing panic in the robust sense. Although the proposed timely disaster alert is effective in a reasonably general environment, it is important to understand its limitations. Below, we discuss these limitations by highlighting the features of the setup that are essential for its effectiveness.

First, notice that the principal wants to dissuade the agents from attacking a fundamentally sound regime, and attacking is irreversible. The same argument cannot be extended to the case where the principal wants to persuade the agents to take an irreversible action. For instance, in a dynamic coordination game in which the agents irreversibly choose when to invest (see Dasgupta 2007), a principal cannot persuade agents to invest by using a disaster alert.[31]

Second, the agents have the option to wait. In some dynamic coordination settings, this may not be the case. The agents may move on different dates, but the dates are exogenously specified. For instance, creditors of a staggered debt portfolio make rollover decisions sequentially. If the agents do not choose the timing of their attack endogenously, then a one-time disclosure may not convince the agents that the other agents who will move later will also do the same. Basak and Zhou (2020) consider such a setup and demonstrate the value of frequent disclosure.[32]

Finally, our principal has a simple objective: she only cares about whether or not the game ends in a disaster. This preference specification is consistent with that of international organizations trying to reduce the panic associated with sovereign debt defaults or of bank supervisors and

[31] One may conjecture that if the attack were also a reversible action, then the result would be stronger because the agents who do not attack would be reassured by the fact that agents who had attacked could change their minds later. However, they may adopt the following strategy ($\mathcal{W}'$): attack right away and reverse the action only if the alert is not triggered. This could trigger the alert, although the disaster was not warranted, and thus, panic may emerge.

[32] In a recent working paper, Koh, Sanguanmoo, and Uzui (2023) consider a dynamic coordination game where agents stochastically get opportunities to revise their decisions. The state can be good or bad, and the agents have homogeneous beliefs over the state. The authors show that the optimal disclosure policy is "conclusive bad news" with a noisy arrival such that upon seeing no bad news, the belief about the state being good keeps moving up.

regulators trying to stop runs on healthy banks. Under such simple preferences, the ex ante commitment is not necessary for our optimal disclosure policy. To see this, note that at disclosure time $\tau$, a disaster alert is triggered when a disaster is inevitable regardless of the agents' actions after time $\tau$. Therefore, the principal has no incentive to misreport or delay the disclosure ex post. Moreover, from an ex ante perspective, policy $\Gamma^\tau$ is able to completely eliminate the panic and obtain the most preferred outcome for the principal. Therefore, she has no incentive to deviate to any other disclosure rules (e.g., send other messages before time $\tau$). However, if the principal has more nuanced preferences, she may need commitment power to implement an EWS or may not want an EWS in the first place.[33] For example, if the principal's objective is to minimize the size of the attack or to delay the attack as much as possible even when a disaster is inevitable (i.e., $\theta \in \Theta^L$), then the timely disaster alert is not a desirable policy. This is because, under this policy, when the disaster is inevitable, all the agents attack and do so immediately after the alert is triggered. In such cases, private disclosure could be useful to ensure miscoordination and therefore delay the crisis (for an interesting example, see Ely 2017).

## Appendix

*Formal Proof of Lemma 1*

*Step 1.*—For any $s_i$, any plan that involves $t_\varnothing \in (0, \tau)$ is strictly dominated by attacking immediately at $t_\varnothing = 0$ ($\mathcal{A}$). Formally, for all $s_i$,

$$(0, t_0, t_1) \succ (t_\varnothing, t_0, t_1) \text{ for all } t_\varnothing \in (0, \tau),\ t_0, t_1 \in [\tau, T] \cup \{\infty\}.$$

To prove this, note that, for all $s_i \in \mathbb{S}$, $t_\varnothing \in (0, \tau)$, and $t_0, t_1 \in [\tau, T] \cup \{\infty\}$,

$$\begin{aligned}
&\mathbb{U}((t_\varnothing, t_0, t_1), \chi_{-i}; s_i) = \mathbb{E}[u(t_\varnothing, \theta, \mathbf{N}(\chi_{-i}))|s_i] \\
&= \mathbb{P}(\theta \in \Theta^L|s_i)\mathbb{E}[u(t_\varnothing, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta^L] \\
&\quad + \mathbb{P}(\theta \in \Theta\backslash\Theta^L|s_i)\mathbb{E}[u(t_\varnothing, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta\backslash\Theta^L] \\
&< \mathbb{P}(\theta \in \Theta^L|s_i)\mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta^L] \\
&\quad + \mathbb{P}(\theta \in \Theta\backslash\Theta^L|s_i)\mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta\backslash\Theta^L] \\
&= \mathbb{U}((0, t_0, t_1), \chi_{-i}; s_i).
\end{aligned}$$

Note that the above inequality holds for any $\chi_{-i}$. This inequality holds because (1) $\mathbb{P}(\theta \in \Theta^L|s_i) > 0$ under assumption 2(A); and (2) according to

[33] Note that, if the principal gets a higher payoff when the size of the attack on a fundamentally sound regime ($\theta \notin \Theta^L$) is smaller, a timely disaster alert remains optimal. This is because, under a timely disaster alert, the size of the attack is 0 when $\theta \notin \Theta^L$. Moreover, if the principal wants to maximize the welfare of the agents, then she should schedule disclosure time $\tau$ as early as possible (but still after time 0) because the agents bear an unnecessary cost while waiting for the alert.

assumption 1(A), $u(t_\varnothing, \theta, \mathbf{N}(\chi_{-i})) < u(0, \theta, \mathbf{N}(\chi_{-i}))$ when $\theta \in \Theta^L$ and $u(t_\varnothing, \theta, \mathbf{N}(\chi_{-i})) \leq u(0, \theta, \mathbf{N}(\chi_{-i}))$ when $\theta \in \Theta \setminus \Theta^L$.

*Step 2.*—For all $s_i$, a plan that involves waiting until the disclosure date but not attacking right away on $d^\tau = 1$ is strictly dominated by a plan that involves waiting until the disclosure date and attacking right away at $\tau$ conditional on $d^\tau = 1$. Formally, for all $s_i$,

$$(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, t_1) \text{ for all } t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}, \text{ and } t_1 \in (\tau, T] \cup \{\infty\}.$$

To see this, consider $t_1 = \infty$ and compare it with $t_1 = \tau$ for any fixed $t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}$. We can prove that for any $t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}$, we have $(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, \infty)$ for any $s_i$ and $\chi_{-i}$. This is because

$$\begin{aligned}
&\mathbb{U}((t_\varnothing, t_0, \infty), \chi_{-i}; s_i) \\
=\ &\mathbb{P}(d^\tau = 1|s_i)\mathbb{E}[u(\infty, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau = 1] \\
&+ \mathbb{P}(d^\tau = 0|s_i)\mathbb{E}[u(t_0, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau = 0] \\
<\ &\mathbb{P}(d^\tau = 1|s_i)\mathbb{E}[u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau = 1] \\
&+ \mathbb{P}(d^\tau = 0|s_i)\mathbb{E}[u(t_0, \theta, \mathbf{N}(\chi_{-i}))|s_i, d^\tau = 0] \\
=\ &\mathbb{U}((t_\varnothing, t_0, \tau), \chi_{-i}; s_i).
\end{aligned}$$

The inequality holds regardless of $s_i$ and $\chi_{-i}$. This is because (1) $\mathbb{P}(d^\tau = 1|s_i) \geq \mathbb{P}(\theta \in \Theta^L|s_i) > 0$ under assumption 2(A); and (2) according to condition (1), if $d^\tau = 1$, $u(\infty, \theta, \mathbf{N}(\chi_{-i})) < u(\tau, \theta, \mathbf{N}(\chi_{-i}))$. Thus, we have $(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, \infty)$.

Next, consider any $t_1 \in (\tau, T]$ and compare it with $t_1 = \tau$ for any fixed $t_\varnothing, t_0 \in [\tau, T] \cup \{\infty\}$. We have

$$\begin{aligned}
&\mathbb{U}((t_\varnothing, t_0, t_1), \chi_{-i}; s_i) \\
=\ &\mathbb{P}(\theta \in \Theta^L|s_i)\mathbb{E}[u(t_1, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta^L] \\
&+ \mathbb{P}(\theta \in \Theta\setminus\Theta^L, d^\tau = 1|s_i)\mathbb{E}[u(t_1, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta\setminus\Theta^L, d^\tau = 1] \\
&+ \mathbb{P}(\theta \in \Theta\setminus\Theta^L, d^\tau = 0|s_i)\mathbb{E}[u(t_0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta\setminus\Theta^L, d^\tau = 0] \\
<\ &\mathbb{P}(\theta \in \Theta^L|s_i)\mathbb{E}[u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta^L] \\
&+ \mathbb{P}(\theta \in \Theta\setminus\Theta^L, d^\tau = 1|s_i)\mathbb{E}[u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta\setminus\Theta^L, d^\tau = 1] \\
&+ \mathbb{P}(\theta \in \Theta\setminus\Theta^L, d^\tau = 0|s_i)\mathbb{E}[u(t_0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta\setminus\Theta^L, d^\tau = 0] \\
=\ &\mathbb{U}((t_\varnothing, t_0, \tau), \chi_{-i}; s_i).
\end{aligned}$$

This inequality holds for any $s_i$ and $\chi_{-i}$. This is because (1) $\mathbb{P}(\theta \in \Theta^L|s_i) > 0$ under assumption 2(A); and (2) according to assumption 1(A), $u(t_1, \theta, \mathbf{N}(\chi_{-i})) < u(\tau, \theta, \mathbf{N}(\chi_{-i}))$ when $\theta \in \Theta^L$ and $u(t_1, \theta, \mathbf{N}(\chi_{-i})) \leq u(\tau, \theta, \mathbf{N}(\chi_{-i}))$ when $\theta \in \Theta \setminus \Theta^L$. Therefore, for all $s_i$, $(t_\varnothing, t_0, \tau) \succ (t_\varnothing, t_0, t_1)$ for all $t_1 \in (\tau, T] \cup \{\infty\}$.

*Step 3.*—Any plan that involves waiting for the disaster alert, attacking at date $t_1 = \tau$ following $d^\tau = 1$, and attacking at some date $t_0 \in [\tau, T]$ following $d^\tau = 0$ is strictly dominated by attacking immediately at $t_\varnothing = 0$ (or $\mathcal{A}$). Formally, for all $s_i$,

$$(0, t_0', t_1') \succ (t_\varnothing, t_0, \tau) \text{ for all } t_0', t_1', t_\varnothing \in [\tau, T] \cup \{\infty\}, \text{ and } t_0 \in [\tau, T].$$

To show this, note that, for all $s_i \in \mathbb{S}$, $t_0'$, $t_1'$, $t_\varnothing \in [\tau, T] \cup \{\infty\}$, and $t_0 \in [\tau, T]$, we have

$$\begin{aligned}
&\mathbb{U}((t_\varnothing, t_0, \tau), \chi_{-i}; s_i) \\
=\ &\mathbb{P}(\theta \in \Theta^L | s_i)\mathbb{E}[u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta^L] \\
&+ \mathbb{P}(\theta \in \Theta \backslash \Theta^L, d^\tau = 1|s_i)\mathbb{E}[u(\tau, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta \backslash \Theta^L, d^\tau = 1] \\
&+ \mathbb{P}(\theta \in \Theta \backslash \Theta^L, d^\tau = 0|s_i)\mathbb{E}[u(t_0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta \backslash \Theta^L, d^\tau = 0] \\
<\ &\mathbb{P}(\theta \in \Theta^L | s_i)\mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta^L] \\
&+ \mathbb{P}(\theta \in \Theta \backslash \Theta^L, d^\tau = 1|s_i)\mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta \backslash \Theta^L, d^\tau = 1] \\
&+ \mathbb{P}(\theta \in \Theta \backslash \Theta^L, d^\tau = 0|s_i)\mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i}))|s_i, \theta \in \Theta \backslash \Theta^L, d^\tau = 0] \\
=\ &\mathbb{E}[u(0, \theta, \mathbf{N}(\chi_{-i}))] = \mathbb{U}((0, t_0', t_1'), \chi_{-i}; s_i).
\end{aligned}$$

The inequality holds for any $\chi_{-i}$. This is so because (1) $\mathbb{P}(\theta \in \Theta^L | s_i) > 0$ under assumption 2(A); and (2) according to assumption 1(A), $u(\tau, \theta, \mathbf{N}(\chi_{-i})) < u(0, \theta, \mathbf{N}(\chi_{-i}))$ when $\theta \in \Theta^L$ and $u(\tau, \theta, \mathbf{N}(\chi_{-i})) \leq u(0, \theta, \mathbf{N}(\chi_{-i}))$ when $\theta \in \Theta \backslash \Theta^L$ and $u(t_0, \theta, \mathbf{N}(\chi_{-i})) \leq u(0, \theta, \mathbf{N}(\chi_{-i}))$. Therefore, for all $s_i$, $(0, t_0', t_1') \succ (t_\varnothing, t_0, \tau)$ for all $t_0 \in [\tau, T]$, $t_0'$, $t_1'$, $t_\varnothing \in [\tau, T] \cup \{\infty\}$. QED

*Proof of Corollary 1*

Note that for any $t \in [0, T]$, the payoff satisfies condition (1)—if $\theta \leq N(T)$, then $u(\infty, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) = 0 - e^{-rt} < 0$—and condition (2)—if $\theta > N(T)$, then $u(\infty, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) = 2 - e^{-rt} > 0$ (condition [3] holds). Moreover, since $2 - e^{-rt} \geq 1$ for any $t \geq 0$, we have a lower bound $g = 1$ to the benefit of not attacking when the country does not default. Delayed attack is costly since $r > 0$ (assumption 1[A] holds) and

$$u(0, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) = 1 - e^{-rt} =: c_1(t).$$

Since $c(0) = 0$ and $c(\cdot)$ is continuously increasing, assumption 1(B) is satisfied. Thus, the corollary follows from theorem 1. Following (9), we have

$$\hat{\tau}_1 = \min\left\{c_1^{-1}\left(\frac{\epsilon}{1-\epsilon} \times g\right), T\right\} = \min\left\{-\frac{1}{r}\ln\left(1 - \frac{\epsilon}{1-\epsilon}\right), T\right\}.$$

QED

*Proof of Corollary 2*

With the stochastic arrival of the adverse shock, the payoff from attacking is now stochastic based on the realization of $t_s$. Note that, if the attacking time $t \leq t_R$, then, by definition of $t_f$, it must be $t \leq t_f$ (regardless of $t_s$) and the payoff from attacking is $\bar{U}$. Otherwise, if the attacking time $t > t_R$, then if the shock arrives at $t_s < t$, the agent attacks after the bank failure (i.e., $t_f < t$) and will get $\underline{U}$; if $t_s \geq t$, then $t_f \geq t$, and the agent will get $\bar{U}$. Therefore, the expected payoff from attacking is

$$u(t,\theta,\mathbf{N}) = \mathbb{1}(t \le t_R(\theta,\mathbf{N}))\bar{U} + \mathbb{1}(t > t_R(\theta,\mathbf{N}))\Big(\mathbb{P}(t_s \ge t)\bar{U} + \mathbb{P}(t_s < t)\underline{U}\Big)$$

$$= \mathbb{1}(t \le t_R(\theta,\mathbf{N}))\bar{U} + \mathbb{1}(t > t_R(\theta,\mathbf{N}))\Big((1 - \lim_{t' \uparrow t} G(t'))\bar{U} + \lim_{t' \uparrow t} G(t')\underline{U}\Big).$$

Since $G(\cdot)$ is continuous and the payoff parameters are bounded, we can replace $\lim_{t' \uparrow t} G(t')$ with $G(t)$. Therefore,

$$u(t,\theta,\mathbf{N}) = \mathbb{1}(t \le t_R(\theta,\mathbf{N}))\bar{U} + \mathbb{1}(t > t_R(\theta,\mathbf{N}))\Big((1 - G(t))\bar{U} + G(t)\underline{U}\Big). \quad \text{(A1)}$$

Moreover, since bank survives in the end (i.e., $t_f(\theta,\mathbf{N},t_s) = \infty$) if, and only if, $t_R(\theta,\mathbf{N}) = \infty$, the payoff from not attacking is

$$u(\infty,\theta,\mathbf{N}) = \mathbb{1}(t_R(\theta,\mathbf{N}) = \infty)\bar{V} + \mathbb{1}(t_R(\theta,\mathbf{N}) < \infty)\underline{V}.$$

Note that, if $R(\theta, N(T)) \le 0$—that is, the game ends in a disaster ($t_R \in [0,T]$)—then, for any $t \in [0,T)$,

$$u(t,\theta,\mathbf{N}) - u(\infty,\theta,\mathbf{N}) = G(t)\underline{U} + (1 - G(t))\bar{U} - \underline{V} > \underline{U} - \underline{V} \ge 0, \quad \text{(A2)}$$

confirming that this payoff specification satisfies condition (1).[34] On the other hand, if $R(\theta, N(T)) > 0$, that is, the game does not end in a disaster ($t_R = \infty$), then for any $t \in [0,T]$, $u(\infty,\theta,\mathbf{N}) - u(t,\theta,\mathbf{N}) = \bar{V} - \bar{U} > 0$, confirming that this payoff specification satisfies condition (2). Moreover, there is a lower bound $g = \bar{V} - \bar{U}$ to the benefit of not attacking when there is no disaster (condition [3] holds).

From (A1), we can see that $u(t,\theta,\mathbf{N})$ is weakly decreasing in $t \in [0,T]$ for any $(\theta,\mathbf{N})$. Additionally, for any possible $\mathbf{N}$, $t_R = 0$ for any $\theta \in \Theta^L$. Therefore, given $\theta \in \Theta^L$, the expected payoff from attacking at time $t \in [0,T]$ is $u(t,\theta,\mathbf{N}) = \mathbb{1}(t = 0)\bar{U} + \mathbb{1}(t > 0)((1 - G(t))\bar{U} + G(t)\underline{U})$, which is strictly decreasing in $t \in [0,T]$ given that $G(\cdot)$ is strictly increasing, thereby confirming this payoff specification satisfies assumption 1(A). Furthermore, for any $(\theta,\mathbf{N})$ and any $t \in (0,T]$,

$$\begin{aligned} u(0,\theta,\mathbf{N}) - u(t,\theta,\mathbf{N}) &= [1 - \mathbb{1}(t \le t_R(\theta,\mathbf{N})) - \mathbb{1}(t > t_R(\theta,\mathbf{N}))(1 - G(t))]\bar{U} \\ &\quad - \mathbb{1}(t > t_R(\theta,\mathbf{N}))G(t)\underline{U} \\ &\le G(t)(\bar{U} - \underline{U}) \coloneqq c_2(t). \end{aligned} \quad \text{(A3)}$$

Since $G(0) = 0$, we have $c(0) = 0$ and $c(\cdot)$ is continuously increasing. Thus, assumption 1(B) holds in this bank run application. Therefore, the corollary follows

[34] In the special case where $\underline{U} = \underline{V}$ (i.e., attacking after the bank failure yields the same payoff as not attacking), condition (1) is violated, but only when the attacking time is set at $t = T$. This is because, if the game ends in a disaster and $t_R = 0$, attacking at any time $t \in [0,T]$ yields a higher payoff of $\bar{U} > \underline{V}$ if the shock arrives after time $t$, which occurs with probability $\mathbb{P}(t_s \ge t) = 1 - G(t)$ (see [A2]). It is important to note that condition (1) is utilized only in step 2 of lemma 1 within the proof of theorem 1. In this context, condition (1) is applied when comparing the decision to attack at time $\tau$ following $d^\tau = 1$ with the decision not to attack following $d^\tau = 1$. Since $\tau \in (0,T)$ by definition, the fact that condition (1) holds for $t \in [0,T)$ is sufficient for this proof and, consequently, for the proof of theorem 1.

from theorem 1. Finally, following (9) and plugging in $g = \bar{V} - \bar{U}$, we can solve $\hat{\tau}_2$ from

$$G(\hat{\tau}_2) = \min\left\{\frac{\epsilon}{1-\epsilon} \times \frac{\bar{V}-\bar{U}}{\bar{U}-\underline{U}}, 1\right\}.$$

QED

*Proof of Corollary 3*

The payoff specification exhibits complementarity. If the game ends in a disaster, $t_f^M \in [0, T + l]$, and for any $t \in [0, T]$, $u(\infty, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) \leq \underline{V} - e^{-r(T+l)}\underline{U} < 0$ (condition [1] holds). On the other hand, if the game does not end in a disaster (i.e., $R(\theta, N(T)) > 0$), $t_f^M = \infty$, and for any $t \in [0, T]$, $u(\infty, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) = \bar{V} - \bar{U}\int_t^{t+l} e^{-rz}\mathrm{d}B((z-t)/l) \geq \bar{V} - \bar{U} > 0$ (condition [2] holds). Note that this gives us the lower bound $g = \bar{V} - \bar{U} > 0$ to the benefit of not attacking when there is no disaster (condition [3] holds).

The payoff specification satisfies assumption 1(A) since, for any $t_f^M(\theta, \mathbf{N})$, a delayed attack is strictly costly. This is because $\bar{U} > \underline{U} > 0$ and there is discounting. Next, we show that for any $t \in (0, l)$, the payoff specification satisfies assumption 1(B). Consider any $t \in (0, l)$, for any $(\theta, \mathbf{N})$, and accordingly, any $t_f^M(\theta, \mathbf{N})$,

$$\begin{aligned}
&u(0, \theta, \mathbf{N}) - u(t, \theta, \mathbf{N}) \\
&= \int_0^l e^{-rz}\left[\mathbb{1}(z \leq t_f^M)\bar{U} + \mathbb{1}(z > t_f^M)\underline{U}\right]\mathrm{d}B\left(\frac{z}{l}\right) - \int_t^{t+l} e^{-rz}\left[\mathbb{1}(z \leq t_f^M)\bar{U} + \mathbb{1}(z > t_f^M)\underline{U}\right]\mathrm{d}B\left(\frac{z-t}{l}\right) \\
&= \int_0^t e^{-rz}\left[\mathbb{1}(z \leq t_f^M)\bar{U} + \mathbb{1}(z > t_f^M)\underline{U}\right]\mathrm{d}B\left(\frac{z}{l}\right) - \int_l^{t+l} e^{-rz}\left[\mathbb{1}(z \leq t_f^M)\bar{U} + \mathbb{1}(z > t_f^M)\underline{U}\right]\mathrm{d}B\left(\frac{z-t}{l}\right) \\
&\quad + \int_t^l e^{-rz}\left[\mathbb{1}(z \leq t_f^M)\bar{U} + \mathbb{1}(z > t_f^M)\underline{U}\right]\left(\mathrm{d}B\left(\frac{z}{l}\right) - \mathrm{d}B\left(\frac{z-t}{l}\right)\right) \\
&\leq \bar{U}\left(\int_0^t e^{-rz}\mathrm{d}B\left(\frac{z}{l}\right) + \int_t^l e^{-rz}\frac{1}{l}\left(b\left(\frac{z}{l}\right) - b\left(\frac{z-t}{l}\right)\right)\mathrm{d}z\right) \leq \bar{U}\left(\int_0^t \mathrm{d}B\left(\frac{z}{l}\right) + \int_t^l \frac{1}{l}\frac{\kappa t}{l}\mathrm{d}z\right) \\
&\leq \bar{U}\left(B\left(\frac{t}{l}\right) + \frac{\kappa t}{l}\right) := c_3(t).
\end{aligned}$$

Note that the above delay cost function $c_3(t)$ is continuously increasing in $t$ and converges to 0 as $t \to 0$, thereby satisfying assumption 1(B). As a result, we have

$$\hat{\tau}_3 = \min\left\{c_3^{-1}\left(\frac{\epsilon}{1-\epsilon}(\bar{V}-\bar{U})\right), l, T\right\}.$$

QED

*Proof of Proposition 1*

We first prove the results under the disaster alert policy $\Gamma^\tau$ and then the ones under no disclosure.

*Policy* $\Gamma^\tau$.—Define $Z(s, \theta) \equiv \theta - \Phi((s-\theta)/\sigma)$. By definition, $Z(s, \theta)$ is strictly increasing in $\theta$ and strictly decreasing in $s$. Therefore, we can define $\alpha(s)$ as the unique solution to $Z(s, \alpha(s)) = 0$. Furthermore, define $h(s) \equiv \alpha(s) + \sigma\Phi^{-1}(p^*(\tau))$ and

$h^{n+1}(s) \equiv h(h^n(s))$ for $n = 1, 2, \ldots$,[35] and denote $\bar{s}^0 \equiv \infty$ and $\underline{s}^0 \equiv -\infty$. We first derive the rationalizable strategies that survive $(n+1)$ $(n = 1, 2, \ldots)$ rounds of elimination of strictly dominated strategies under disaster alert policy $\Gamma^\tau$.

LEMMA 4. Under the disaster alert $\Gamma^\tau$ with any $\tau \in (0, T)$, the only strategies that survive $(n+1)$ rounds of elimination of strictly dominated strategies (for $n = 0, 1, \ldots$) satisfy

$$\chi_i(s_i) = \begin{cases} \mathcal{W} & \text{if } s_i > h^n(\bar{s}^0), \\ \Delta(\{\mathcal{W}, \mathcal{A}\}) & \text{if } h^n(\underline{s}^0) \leq s_i \leq h^n(\bar{s}^0), \\ \mathcal{A} & \text{if } s_i < h^n(\underline{s}^0). \end{cases} \tag{A4}$$

*Proof.* Round 1: Note that, for all possible $s_i$, we have $\mathbb{P}(\theta \in \Theta^L = (-\infty, 0] | s_i) > 0$ and $\mathbb{P}(\theta \in \Theta^U = (1, +\infty) | s_i) > 0$. Then, following lemma 1, in the first round of elimination, we can eliminate strategies that choose a plan other than $\{\mathcal{A}, \mathcal{W}\}$ for all $s_i$. Therefore, (A4) holds for $n = 0$.

Next, we prove by induction: if, after the $k$th round of elimination $(k = 1, \ldots)$, (A4) holds true for $n = k-1$, then (A4) holds for $n = k$ after the $(k+1)$th round of elimination. Recall that, since lemma 1 holds true, the regime changes if and only if the alert is triggered at time $\tau$ (lemma 3). Accordingly, the regime change condition is identical to the condition of $d^\tau = 1$, that is, $\theta \leq N(0)$. Note that for an agent with private signal $s_i$, the payoff of $\mathcal{A}$ is $\bar{U}$ and the expected payoff of $\mathcal{W}$ is

$$\begin{aligned} &\mathbb{P}(\theta > N(0) | s_i) \times \bar{V} + \mathbb{P}(\theta \leq N(0) | s_i) \times \left[ \mathbb{P}(t_s \geq \tau) \bar{U} + \mathbb{P}(t_s < \tau) \underline{U} \right] \\ &= \left[ \bar{V} - \left( (1 - G(\tau)) \bar{U} + G(\tau) \underline{U} \right) \right] \mathbb{P}(\theta > N(0) | s_i) + \left( (1 - G(\tau)) \bar{U} + G(\tau) \underline{U} \right). \end{aligned}$$

Thus, an agent with private signal $s_i$ strictly prefers $\mathcal{W}$ to $\mathcal{A}$ if

$$\mathbb{P}(\theta > N(0) | s_i) > \frac{\bar{U} - \left( (1 - G(\tau)) \bar{U} + G(\tau) \underline{U} \right)}{\bar{V} - \left( (1 - G(\tau)) \bar{U} + G(\tau) \underline{U} \right)} = p^*(\tau)$$

and strictly prefers $\mathcal{A}$ to $\mathcal{W}$ if $\mathbb{P}(\theta > N(0) | s_i) < p^*(\tau)$.

Round $(k+1)$: On the upper side, since (A4) holds true for $n = k-1$, an agent with $s_i > \bar{s}^{k-1} \equiv h^{k-1}(\bar{s}^0)$ chooses $\mathcal{W}$ for sure. Therefore, the mass of attacks at time 0 satisfy $N(0) \leq \mathbb{P}(s_i \leq \bar{s}^{k-1} | \theta) = \Phi((\bar{s}^{k-1} - \theta)/\sigma)$. By definition of $Z(s, \theta)$, $\theta - N(0) \geq \theta - \Phi((\bar{s}^{k-1} - \theta)/\sigma) = Z(\bar{s}^{k-1}, \theta)$. As a result, by definition of $\alpha(s)$, if $\theta > \alpha(\bar{s}^{k-1})$, $\theta - N(0) > Z(\bar{s}^{k-1}, \alpha(\bar{s}^{k-1})) = 0$, and there is no disaster alert or regime change with certainty. Thus, regardless of what others play (as long as their strategies satisfy [A4] for $n = k-1$), an agent with $s_i$ believes that $\mathbb{P}(\theta > N(0) | s_i) \geq \mathbb{P}(\theta > \alpha(\bar{s}^{k-1}) | s_i) = \Phi((s_i - \alpha(\bar{s}^{k-1}))/\sigma)$. Accordingly, for any

$$\begin{aligned} s_i > \alpha(\bar{s}^{k-1}) + \sigma \Phi^{-1}(p^*(\tau)) &= \alpha(h^{k-1}(\bar{s}^0)) + \sigma \Phi^{-1}(p^*(\tau)) \\ &= h\left(h^{k-1}(\bar{s}^0)\right) = h^k(\bar{s}^0), \end{aligned}$$

[35] Note that $h^0(x) = x$ and $h^1(x) = h(x)$.

since $\mathbb{P}(\theta > N(0)|s_i) > \Phi((h(\bar{s}^{k-1}) - \alpha(\bar{s}^{k-1}))/\sigma) = p^*(\tau)$, $\mathcal{W}$ generates a strictly higher expected payoff than $\mathcal{A}$ regardless of what (iteratedly) undominated strategies others play. Thus, $\mathcal{A}$ is eliminated for $s_i > h^k(\bar{s}^0)$.

On the lower side, since (A4) holds true for $n = k - 1$, an agent with $s_i < \underline{s}^{k-1} \equiv h^{k-1}(\underline{s}^0)$ chooses $\mathcal{A}$ for sure. Hence, the mass of attacks at time 0 satisfy $N(0) \geq \mathbb{P}(s_i < \underline{s}^{k-1}|\theta) = \Phi((\underline{s}^{k-1} - \theta)/\sigma)$. By definition of $Z(s, \theta)$, $\theta - N(0) \leq \theta - \Phi((\underline{s}^{k-1} - \theta)/\sigma) = Z(\underline{s}^{k-1}, \theta)$. According to the definition of $\alpha(s)$, if $\theta < \alpha(\underline{s}^{k-1})$, $\theta - N(0) < Z(\underline{s}^{k-1}, \alpha(\underline{s}^{k-1})) = 0$, and therefore, the disaster alert and regime change happen with certainty. Hence, regardless of what others play (as long as their strategies satisfy [A4] for $k$), an agent with $s_i$ believes that $\mathbb{P}(\theta > N(0)|s_i) \leq 1 - \mathbb{P}(\theta < \alpha(\underline{s}^{k-1})|s_i) = \Phi((s_i - \alpha(\underline{s}^{k-1}))/\sigma)$. Therefore, for

$$\begin{aligned} s_i < \alpha(\underline{s}^{k-1}) + \sigma\Phi^{-1}(p^*(\tau)) &= \alpha(h^{k-1}(\underline{s}^0)) + \sigma\Phi^{-1}(p^*(\tau)) \\ &= h\left(h^{k-1}(\underline{s}^0)\right) = h^k(\underline{s}^0), \end{aligned}$$

$\mathbb{P}(\theta > N(0)|s_i) < \Phi((h(\underline{s}^{k-1}) - \alpha(\underline{s}^{k-1}))/\sigma) = p^*(\tau)$. This implies that, regardless of what (iteratedly) undominated strategies others play, $\mathcal{A}$ generates a strictly higher expected payoff than $\mathcal{W}$. Hence, $\mathcal{W}$ is eliminated for $s_i < h(\underline{s}^{k-1}) = h^k(\underline{s}^0)$.

Therefore, the strategies that survive a $(k + 1)$ round of elimination of strictly dominated strategies satisfy (A4) for $n = k$. This completes the proof. QED

Since $\alpha(s)$ is increasing in $s$ with $\alpha(\underline{s}^0) = 0$ and $\alpha(\bar{s}^0) = 1$, $h(\bar{s}^0) < \bar{s}^0 = +\infty$, $\{h^n(\bar{s}^0)\}_{n=1}^{\infty}$ is a decreasing sequence in a bounded interval $[\sigma\Phi^{-1}(p^*(\tau)), 1 + \sigma\Phi^{-1}(p^*(\tau))]$. Therefore, $\{h^n(\bar{s}^0)\}_{n=1}^{\infty}$ converges monotonically to $\bar{s}^*$ such that $\bar{s}^* = h(\bar{s}^*)$, which gives $\bar{s}^* = p^*(\tau) + \sigma\Phi^{-1}(p^*(\tau))$. Thus, $\mathcal{A}$ is eliminated for $s_i > \bar{s}^* = p^*(\tau) + \sigma\Phi^{-1}(p^*(\tau))$. Similarly, $\{h^n(\underline{s}^0)\}_{n=1}^{\infty}$ converges monotonically to $\underline{s}^* = p^*(\tau) + \sigma\Phi^{-1}(p^*(\tau))$.

As $\bar{s}^* = \underline{s}^* = p^*(\tau) + \sigma\Phi^{-1}(p^*(\tau)) \equiv s^*$, the unique rationalizable strategy is $\mathcal{A}$ for $s_i \leq s^*$ and $\mathcal{W}$ for $s_i > s^*$. Therefore, the regime survives if and only if $\theta - \Phi((s^* - \theta)/\sigma) > 0$; that is,

$$\theta > p^*(\tau) = \frac{\bar{U} - \left((1 - G(\tau))\bar{U} + G(\tau)\underline{U}\right)}{\bar{V} - \left((1 - G(\tau))\bar{U} + G(\tau)\underline{U}\right)} \equiv \theta^*(\tau).$$

As $G(t)$ is continuously increasing in $t$, $\theta^*(\tau)$ is continuously increasing in $\tau$. Since $\lim_{\tau \downarrow 0}\theta^*(\tau) = 0$, for any $\zeta > 0$, there exists

$$\hat{\tau}(\zeta) = G^{-1}\left(\min\left\{\frac{\bar{V} - \bar{U}}{\bar{U} - \underline{U}}\frac{\zeta}{1 - \zeta}, 1\right\}\right)$$

such that with $\tau \in (0, \hat{\tau}(\zeta))$, $\theta^*(\tau) < \zeta$.

*No disclosure.*—With no disclosure, because waiting is costly (assumption 1[A]), for all $s_i$, the only rationalizable strategies are "attacking immediately at time 0" and "not attacking at all." For an agent with $s_i$, the payoff of attacking immediately at time 0 is $\bar{U}$, and the expected payoff of not attacking is

$$\mathbb{P}(\theta > N(0)|s_i) \times \bar{V} + \mathbb{P}(\theta \leq N(0)|s_i) \times \underline{V} = \underline{V} + (\bar{V} - \underline{V})\,\mathbb{P}(\theta > N(0)|s_i).$$

Thus, an agent with private signal $s_i$ strictly prefers not attacking (attacking immediately) if

$$\mathbb{P}(\theta > N(0)|s_i) > \frac{\bar{U} - \underline{V}}{\bar{V} - \underline{V}} = \hat{p}$$

($\mathbb{P}(\theta > N(0)|s_i) < \hat{p}$). Hence, by replacing $p^*(\tau)$ with $\hat{p}$ in the proof of iterated elimination (starting from round 2) in the case under the $\Gamma^\tau$ policy, we can easily show that the unique rationalizable strategy is to attack at $t = 0$ for $s_i \leq \hat{p} + \sigma\,\Phi^{-1}(\hat{p})$ and never attack for $s_i > \hat{p} + \sigma\,\Phi^{-1}(\hat{p})$. The regime survives if and only if $\theta > \hat{\theta} = \hat{p}$. As for any $\tau \in (0, T)$, $\mathbb{P}(t_s < \tau) = G(\tau) > 0$, we have

$$p^*(\tau) = \frac{\bar{U} - \left((1 - G(\tau))\bar{U} + G(\tau)\underline{U}\right)}{\bar{V} - \left((1 - G(\tau))\bar{U} + G(\tau)\underline{U}\right)} < \frac{\bar{U} - \underline{V}}{\bar{V} - \underline{V}} = \hat{p},$$

so $\theta^*(\tau) < \hat{\theta}$ and the policy $\Gamma^\tau$ reduces the probability of panic. QED

*Proof of Proposition 2*

Under the extended $\Gamma^\tau$ policy, where $\tau < \bar{\tau}$, each agent chooses a contingency plan $(t^k, t_0^k, t_1^k)_{k=0}^K$, where $t^k$, $t_0^k$, and $t_1^k$ denote the time of attack after receiving the $(k+1)$th private signal at time $\nu_k$, receiving $d^{\nu_k+\tau} = 0$ at time $\nu_k + \tau$, and receiving $d^{\nu_k+\tau} = 1$ at time $\nu_k + \tau$, respectively.

From the same argument as in lemma 1, it follows that an agent either attacks at $\nu_k$ or waits at least until the next information arrives at $\nu_k + \tau$, and if he waits, then he must follow the alert. Since the regime will surely change following $d^{\nu_k+\tau} = 1$, agents must attack at $\nu_k + \tau$ after receiving $d^{\nu_k+\tau} = 1$. This implies, after $d^{\nu_k+\tau} = 0$, they must wait at least until the next information arrives at time $\nu_{k+1}$. Thus, any strategy that is not strictly dominated involves choosing some plan among $\{\mathcal{W}^k\}_{k=0}^{K+1}$ for all possible $\mathbf{s}_i$, in which $\mathcal{W}^k$ ($k = 1, 2, \ldots, K$) means that the agent waits for the first $k$ alerts and does not attack unless an alert is triggered but then attacks after receiving the $(k+1)$th private signal (at $\nu_k$) and does not wait for the next alert (at time $\nu_k + \tau$). Technically, the plan $\mathcal{W}^k$ implies

(1) for $l < k$, $t^l \geq \nu_l + \tau$, $t_0^l \geq \nu_{l+1}$, and $t_1^l = \nu_l + \tau$;
(2) for $l = k$, $t^l = \nu_l$, $t_0^l \geq \nu_l + \tau$, and $t_1^l \geq \nu_l + \tau$; and
(3) for $l > k$, $t^l \geq \nu_l$, $t_0^l \geq \nu_l + \tau$, and $t_1^l \geq \nu_l + \tau$.

In addition, $\mathcal{W}^0 = \mathcal{A}$, which means $t^0 = \nu_0 = 0$, $t^l \geq \nu_l$ for all $l > 0$, and $t_0^l$, $t_1^l \geq \nu_l + \tau$ for all $l \geq 0$. The plan $W^{K+1}$ means waiting for all the alerts and never attacking unless an alert is triggered; that is, $t^l \geq \nu_l + \tau$, $t_0^l \geq \nu_{l+1}$, and $t_1^l = \nu_l + \tau$ for all $l \leq K$.

Next, we prove that $\mathcal{W}^{K+1}$ is the unique rationalizable strategy for any type $\mathbf{s}_i$. We prove this by iterated elimination: we first establish the strict dominance of $\mathcal{W}^{K+1}$ over $\mathcal{W}^K$, and then we show that, for any $k \in \{0, 1, \ldots, K-1\}$, if $\mathcal{W}^{K+1}$ strictly dominates any $\mathcal{W}^{k'}$ for all $k' > k$, then $\mathcal{W}^{K+1}$ strictly dominates $\mathcal{W}^k$.

To prove that $\mathcal{W}^K$ is strictly dominated by $\mathcal{W}^{K+1}$, first note that both strategies follow the same path if any of the first $K$ alerts is triggered, or $d^{\nu_{K-1}+\tau} = 1$. Therefore, when comparing these two strategies, we need only consider the histories following $d^{\nu_{K-1}+\tau} = 0$.

Conditional on $d^{\nu_K+\tau} = 0$, under all possible rationalizable strategy profiles $\chi_{-i} = (\chi_j)_{j\neq i}$ in which $\chi_j(\mathbf{s}_j) \in \{\mathcal{W}^k\}_{k=0}^{K+1}$ for all $\mathbf{s}_j$, there is no further attack, which implies $(\theta, \mathbf{N}) \notin \mathcal{D}$. Therefore, similar to (5), the benefit of $\mathcal{W}^{K+1}$ compared with $\mathcal{W}^K$, conditional on $d^{\nu_K+\tau} = 0$, is

$$\begin{aligned}&\mathbb{B}_K(\Gamma^\tau, \chi_{-i}, \mathbf{s}_i)\\ &= \mathbb{E}[u(\infty, \theta, \mathbf{N}(\chi_{-i})) - u(\nu_K, \theta, \mathbf{N}(\chi_{-i}))|\mathbf{s}_i, d^{\nu_K+\tau}(\theta, \mathbf{N}_{\nu_K+\tau}(\chi_{-i})) = 0] \geq g.\end{aligned}$$

Conditional on $d^{\nu_K+\tau} = 1$, under contingency plan $\mathcal{W}^{K+1}$, the agent attacks at time $\nu_K + \tau$, whereas he attacks at $\nu^K$ under contingency plan $\mathcal{W}^K$. Therefore, similar to (7), the expected cost of $\mathcal{W}^{K+1}$ compared with $\mathcal{W}^K$ can be written as

$$\begin{aligned}\mathbb{C}_K(\Gamma^\tau, \chi_{-i}, \mathbf{s}_i) &= \mathbb{E}[u(\nu_K, \theta, \mathbf{N}(\chi_{-i})) - u(\nu_K + \tau, \theta, \mathbf{N}(\chi_{-i}))|\mathbf{s}_i, d^{\nu_K+\tau}(\theta, \mathbf{N}_{\nu_K+\tau}(\chi_{-i})) = 1]\\ &\leq c(\nu_K + \tau) - c(\nu_K).\end{aligned}$$

Therefore, the expected payoff difference between $\mathcal{W}^{K+1}$ and $\mathcal{W}^K$ is

$$\begin{aligned}&\mathbb{D}_K(\Gamma^\tau, \chi_{-i}, \mathbf{s}_i)\\ &= \mathbb{P}(d^{\nu_{K-1}+\tau}(\theta, \mathbf{N}_{\nu_{K-1}+\tau}(\chi_{-i})) = 1|\mathbf{s}_i) \cdot 0\\ &\quad + \mathbb{P}(d^{\nu_{K-1}+\tau}(\theta, \mathbf{N}_{\nu_{K-1}+\tau}(\chi_{-i})) = 0, d^{\nu_K+\tau}(\theta, \mathbf{N}_{\nu_K+\tau}(\chi_{-i})) = 0|\mathbf{s}_i)\mathbb{B}_K(\Gamma^\tau, \chi_{-i}, \mathbf{s}_i)\\ &\quad - \mathbb{P}(d^{\nu_{K-1}+\tau}(\theta, \mathbf{N}_{\nu_{K-1}+\tau}(\chi_{-i})) = 0, d^{\nu_K+\tau}(\theta, \mathbf{N}_{\nu_K+\tau}(\chi_{-i})) = 1|\mathbf{s}_i)\mathbb{C}_K(\Gamma^\tau, x_{-i}, \mathbf{s}_i)\\ &> \epsilon \cdot g - (1-\epsilon) \cdot [c(\nu_K + \tau) - c(\nu_K)].\end{aligned}$$

The last inequality follows from the fact that, under assumption 2(B), for any $\mathbf{s}_i$, $\mathbb{P}(d^{\nu_K+\tau} = d^{\nu_{K-1}+\tau} = 0|\mathbf{s}_i) \geq \mathbb{P}(\theta \in \Theta^U|\mathbf{s}_i) > \epsilon$, then the expected benefit $\mathbb{B}_K(\cdot)$ has a lower bound $g$, whereas the expected cost $\mathbb{C}_K(\cdot)$ has an upper bound $(c(\nu_K + \tau) - c(\nu_K))$. Since $c(t)$ is continuous (i.e., $\lim_{\tau\downarrow 0}(c(\nu_K + \tau) - c(\nu_K)) = 0$), as in lemma 3, for any $\epsilon > 0$, we can find $\tilde{\tau}_K \equiv \min\{c^{-1}(g \times \epsilon/(1-\epsilon) + c(\nu_K)) - \nu_K, T\} > 0$ such that for any $\tau \in (0, \min\{\tilde{\tau}_K, \bar{\tau}\})$, $\mathbb{D}_K(\Gamma^\tau, \chi_{-i}, \mathbf{s}_i) > 0$ irrespective of $\mathbf{s}_i$. As this holds true for any $\chi_{-i}$, $\mathcal{W}^{K+1}$ strictly dominates $\mathcal{W}^K$ for all $\mathbf{s}_i$.

Now, given that $\mathcal{W}^K$ is strictly dominated, there is no attack at time $\nu_K$ under any undominated strategies. Therefore, the second-to-last alert $d^{\nu_{K-1}+\tau}$ perfectly predicts a disaster; that is, $d^{\nu_K+\tau} = d^{\nu_{K-1}+\tau} = \mathbb{1}((\theta, \mathbf{N}) \in \mathcal{D})$. Next, we repeat the same argument and evaluate the expected benefit and cost of strategy $\mathcal{W}^{K+1}$ compared with strategy $\mathcal{W}^{K-1}$. There is no difference between these two strategies if $d^{\nu_{K-2}+\tau} = 1$. Conditional on $d^{\nu_{K-2}+\tau} = 0$, the benefit of $\mathcal{W}^{K+1}$ comes from the event when $d^{\nu_{K-1}+\tau} = 0$ (and accordingly no disaster), whereas the cost comes from the event when $d^{\nu_{K-1}+\tau} = 1$ (and accordingly a disaster). Conditional on $d^{\nu_{K-2}+\tau} = 0$ and $d^{\nu_{K-1}+\tau} = 1$, the agent attacks at time $\nu_{K-1}$ under $\mathcal{W}^{K-1}$, and he attacks at time

$\nu_{K-1} + \tau$ under $\mathcal{W}^{K+1}$. Following the same procedures as we did for the comparison between $\mathcal{W}^{K+1}$ and $\mathcal{W}^{K}$, we can show the existence of the cutoff $\tilde{\tau}_{K-1} \equiv \min\{c^{-1}(g \times \epsilon/(1-\epsilon) + c(\nu_{K-1})) - \nu_{K-1}, T\} > 0$ and establish the strict dominance of $\mathcal{W}^{K+1}$ over $\mathcal{W}^{K-1}$ under any policy $\Gamma^{\tau}$ with $\tau \in (0, \min\{\tilde{\tau}_{K-1}, \tilde{\tau}_{K}, \bar{\tau}\})$.

Following the same procedures, we can find $(K+1)$ cutoffs $\{\tilde{\tau}_k\}_{k=0}^{K}$ and establish the strict dominance of $\mathcal{W}^{K+1}$ over $\mathcal{W}^{k}$ (for all $k = 0, 1, \ldots, K$) under any disclosure policy $\Gamma^{\tau}$ with $\tau \in (0, \min\{\tilde{\tau}, \bar{\tau}\})$, where $\tilde{\tau} \coloneqq \min_{k=0}^{K}\tilde{\tau}_k > 0$. This completes the proof. QED

## References


Anderson, R. W. 2016. *Stress Testing and Macroprudential Regulation: A Transatlantic Assessment.* Paris: CEPR.

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan. 2007. "Dynamic Global Games of Regime Change: Learning, Multiplicity, and the Timing of Attacks." *Econometrica* 75 (3): 711–56.

Angeletos, George-Marios, and Chen Lian. 2017. "Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination." In *Handbook of Macroeconomics*, vol. 2, 1065–240. North-Holland: Elsevier.

Au, Pak Hung. 2015. "Dynamic Information Disclosure." *RAND J. Econ.* 46 (4): 791–823.

Ball, Ian. 2023. "Dynamic Information Provision: Rewarding the Past and Guiding the Future." *J. Econometric Soc.*, https://doi.org/10.3982/ECTA17345.

Basak, Deepal, and Zhen Zhou. 2020. "Diffusing Coordination Risk." *A.E.R.* 110 (1): 271–97.

Ben-Porath, Elchanan, and Eddie Dekel. 1992. "Signaling Future Actions and Potential for Sacrifice." *J. Econ. Theory* 57:36–51.

Bergemann, Dirk, and Stephen Morris. 2019. "Information Design: A Unified Perspective." *J. Econ. Literature* 57 (1): 44–95.

Bernanke, Ben S. 2013. "Stress Testing Banks: What Have We Learned?"

Chamley, Christophe. 1999. "Coordinating Regime Switches." *Q.J.E.* 114 (3): 869–905.

———. 2003. "Dynamic Speculative Attacks." *A.E.R.* 93 (3): 603–21.

Chamley, Christophe, and Douglas Gale. 1994. "Information Revelation and Strategic Delay in a Model of Investment." *Econometrica* 62 (5): 1065–85.

Chassang, Sylvain. 2010. "Fear of Miscoordination and the Robustness of Cooperation in Dynamic Global Games with Exit." *Econometrica* 78 (3): 973–1006.

Dasgupta, Amil. 2007. "Coordination and Delay in Global Games." *J. Econ. Theory* 134 (1): 195–225.

Ely, Jeffrey C. 2017. "Beeps." *A.E.R.* 107 (1): 31–53.

Ely, Jeffrey C., George Georgiadis, Sina Moghadas Khorasani, and Luis Rayo. 2023. "Optimal Feedback in Contests." *Rev. Econ. Studies* 90 (5): 2370–94.

Ely, Jeffrey C., and Martin Szydlowski. 2020. "Moving the Goalposts." *J.P.E.* 128 (2): 468–506.

Goldstein, Itay, and Chong Huang. 2016. "Bayesian Persuasion in Coordination Games." *A.E.R.* 106 (5): 592–96.

Goldstein, Itay, and Yaron Leitner. 2018. "Stress Tests and Information Disclosure." *J. Econ. Theory* 177:34–69.

Gorton, Gary. 2015. "Stress for Success: A Review of Timothy Geithner's Financial Crisis Memoir." *J. Econ. Literature* 53 (4): 975–95.

Inostroza, Nicolas. 2019. "Persuading Multiple Audiences: An Information Design Approach to Banking Regulation." Working paper, https://doi.org/10.2139/ssrn.3450981.

Inostroza, Nicolas, and Alessandro Pavan. 2025. "Adversarial Coordination and Public Information Design." *Theoretical Econ.*, forthcoming.

Kamenica, Emir. 2019. "Bayesian Persuasion and Information Design." *Ann. Rev. Econ.* 11: 249–72.

Kamenica, Emir, and Matthew Gentzkow. 2011. "Bayesian Persuasion." *A.E.R.* 101 (6): 2590–615.

Koh, Andrew, Sivakorn Sanguanmoo, and Kei Uzui. 2023. "Full Implementation with Dynamic Information." Working paper, https://doi.org/10.2139/ssrn.4669928.

Kohlberg, Elon, and Jean-Francois Mertens. 1986. "On the Strategic Stability of Equilibria." *Econometrica* 54 (5): 1003–37.

Li, Fei, Yangbo Song, and Mofei Zhao. 2023. "Global Manipulation by Local Obfuscation." *J. Econ. Theory* 207:105575.

Li, Xuelin, Martin Szydlowski, and Fangyuan Yu. 2025. "Dynamic Information Design in an Entry Game." *J. Econ. Theory*.

Mathevet, Laurent, Jacopo Perego, and Ina Taneva. 2020. "On Information Design in Games." *J.P.E.* 128 (4): 1370–404.

Moore, John, and Rafael Repullo. 1988. "Subgame Perfect Implementation." *Econometrica* 56 (5): 1191–220.

Morris, Stephen, Daisuke Oyama, and Satoru Takahashi. 2024. "Implementation via Information Design in Binary-Action Supermodular Games." *Econometrica* 92 (3): 775–813.

Morris, Stephen, and Hyun Song Shin. 2003. "Global Games: Theory and Applications." In *Advances in Economics and Econometrics (Proceeding of the Eighth World Congress of the Econometric Society)*, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky. Cambridge, MA: Cambridge University Press.

Orlov, Dmitry, Andrzej Skrzypacz, and Pavel Zryumov. 2020. "Persuading the Principal to Wait." *J.P.E.* 128 (7): 2542–78.

Orlov, Dmitry, Pavel Zryumov, and Andrzej Skrzypacz. 2023. "Design of Macro-Prudential Stress Tests." *Rev. Financial Studies* 36 (11): 4460–501.

Peristiani, Stavros, Donald P. Morgan, and Vanessa Savino. 2010. "The Information Value of the Stress Test and Bank Opacity." Staff Report no. 460, Fed. Res. Bank, New York.