Timely Persuasion

Deepal Basak, Zhen Zhou*

January 2020

Abstract

We propose a simple dynamic information disclosure policy that eliminates panic. Panic occurs when some agents take an undesirable action (attack) because they fear that other agents will behave similarly, and thus causing regime change even though it is unwarranted. We consider a mass of privately informed agents who can attack a regime at any point within a time window. The attack is irreversible, waiting is costly, and the waiting cost is continuous. The policy we propose is called "disaster alert," which, at a given time, publicly discloses whether the regime is going to change regardless of the future actions of agents. We show that a timely alert persuades agents to wait for the alert and not attack if the alert is not triggered, regardless of their private information, thus eliminating panic. We apply this theory to demonstrate how forward-looking stress tests can help stop bank runs.

JEL Classification Numbers: *D02, D82, D83, G28*
Key Words: *Coordination, Information Design, Panic*

*We thank Emiliano Catonini, Christophe Chamley, Joyee Deb, Douglas Gale, Ju Hu, Nicolas Inostroza, Stephen Morris, Alessandro Pavan, David Pearce, Antonio Penta, Ennio Stacchetti, Colin Stewart, Debraj Ray, Laura Veldkamp, and Xingye Wu as well as the seminar participants at the Econometric Society meetings, Stony Brook, Peking University, Tsinghua University, Fudan, NYU Shanghai, NUS, UPF, SMU, University of Washington, Toronto, Kansas, Indiana Kelley, Rochester Simon, UIUC, Utah Eccles for their helpful comments and suggestions. We thank Samyak Jain and Dandan Zhao for their excellent research assistance. Basak: Indian School of Business, Email: deepal_basak@isb.edu; Zhou: PBC School of Finance, Tsinghua University, Email: zhouzh@pbcsf.tsinghua.edu.cn

Introduction

In a game of strategic complementarity, often, an agent takes an undesirable action because he expects other agents to behave similarly, even when such an action is not warranted. We call this *panic*. Suppose some investors have made direct investments in an emerging market, and an adverse shock hits the emerging economy. These investors have noisy information on the fundamental strength of the economy to face such a shock. If the fundamental strength is extremely poor, investors should exit the market. However, even if it is not extremely poor, because of strategic complementarity, investors may panic and begin to exit the market if they believe other investors will also exit. Can such panic be avoided? We propose a simple dynamic information disclosure policy that eliminates panic.

The above example can be appropriately captured through a canonical regime change game: A mass of agents decides whether to attack a regime. If the regime is strong enough to withstand the aggregate attack, the regime survives; otherwise, the regime changes (or the status quo fails). Also, we consider a principal or an information designer who wants the regime to survive.

The literature usually assumes that agents move simultaneously. However, in practice, attacks do not occur in one instance. An information designer may selectively disclose information about the history of attacks to dissuade agents from attacking. Also, the agents may wait for new information from the information designer before they attack. To this end, we extend the canonical regime change game and allow agents to attack within a time window $[0, T]$. We assume that attacking is an irreversible action and waiting is costly.

In our benchmark setup, we assume that the fate of the regime is decided at the end of the time window depending on the underlying strength of the regime and the aggregate attack. Let us reconsider the capital outflow example. It is reasonable to assume that once an investor exits the market, he does not come back. If an investor chooses to wait and exit later, he may lose the interest income from investing elsewhere. In this sense, waiting is costly. It takes time for investors to move capital away from an emerging economy. He may start talking to real estate agents or firing employees. In such a slow-moving capital market, the regime does not change instantaneously. Since the regime does not change in the middle of the time window, it is reasonable to assume that the delay cost is continuous in time.[1]

[1] We relax this assumption later when we consider financial markets. In a financial market, withdrawals could be immediate, and thus, the regime could change in the middle of the time window.

The agents are uncertain about the fundamental strength of the regime; and before the game begins, they receive some noisy private signals. Unlike the global game literature, we do not impose independence on the noise. The noises may be independent or correlated. This allows for some of the agents to share their private information as well, and thus, the beliefs could be homogeneous or arbitrarily heterogeneous. Based on these signals, agents form their beliefs about the fundamental and other agents' signals. For simplicity, let us suppose that the agents receive no additional information over time. If an agent believes that the regime is unlikely to survive, he attacks. If the regime is fundamentally strong, or if many of the other agents act otherwise, then the regime may still survive. Thus, attacking immediately could be a mistake ex-post. We assume that regardless of his signal and other agents' actions, an agent always assigns a strictly positive probability to his action being a mistake. We refer to this assumption as *doubt*.[2]

The principal knows the state or the fundamental strength, and the aggregate attack at any time. She commits to a dynamic information disclosure rule: At some time t, she will send messages to the agents based on the exogenous fundamental and the endogenous history of the attack until time t. Thus, unlike most of the literature on information design that only focuses on disclosure about an exogenous state, we also explore disclosure about the endogenous history of attacks.

Let us first consider two extreme disclosure policies: no disclosure and full disclosure. If the principal does not disclose any information, then the game is equivalent to a static regime change game. It is well known that, under sufficiently heterogenous beliefs, the iterated elimination of never-best responses leads to a unique cutoff strategy — the agents attack if and only if their private signals are below some threshold. Under such a cut-off strategy, a regime fails due to panic when the fundamental is neither weak enough to warrant an attack nor strong enough to withstand the aggregate attack.

On the other hand, if the principal adopts a full disclosure policy, then it is a complete information regime change game. Under complete information, if the fundamental is not good enough for the regime to survive regardless of aggregate attack, then there is always an equilibrium in which all agents attack.

Since there could be multiple equilibria under any given disclosure policy, we focus on the rationalizable set of strategies.[3] We formally define panic as the set of states in which a

[2]This assumption is trivially true under the standard global game information structure when the noise distribution has full support. For our result, it is sufficient to assume only that the agent doubts that attacking could be a mistake. This is not a necessary assumption, and we discuss this in section 5.

[3]Notably, we assume common belief in sequential rationality at the beginning of the game. Unlike the

regime could fail when some agents attack and attack is a rationalizable strategy for them, but the regime will not fail if the agents do not attack. The principal chooses a policy so as to minimize the chance of panic.

We propose a simple partial disclosure policy called the "disaster alert." A disaster alert at a pre-specified time t is triggered if the regime is no longer strong enough to withstand any future attacks. Thus, a disaster alert at a pre-specified date publicly sends an early warning to the agents if it becomes evident that the regime cannot survive.

When the disaster alert is triggered, the dominant strategy is to attack. On the other hand, if the disaster alert is not triggered, then the agents only learn that attacking is not the dominant strategy. However, if the other agents attack after seeing no alert, then the regime may fail. Thus, it is wise for an agent to attack if he anticipates that many of the other agents will attack even when the alert is not triggered. Interestingly, if a sequentially rational agent has waited for the disaster alert, it must hold that he will not attack when the alert is not triggered; otherwise, there is no positive option value of waiting to justify the cost of a delayed attack.

Thus, any agent who has waited for the disclosure of disaster alert will attack afterwards if and only if the alert is triggered. In other words, all these agents who have waited follow the alert and coordinate their actions perfectly. Consequently, if they believe that the other agents are sequentially rational, then once the principal discloses the disaster news, there is no strategic uncertainty left – if the alert is triggered, all these agents attack and the regime fails; otherwise, none of them attack and the regime survives.

However, this does not mean that the agents will wait for the alert. An agent who receives a low signal may think that the disaster alert is very likely to be triggered, and hence attack than wait for the alert. If some of the agents attack, then it may trigger the disaster alert, even though the fundamental does not warrant it.

Recall that an agent always doubts whether attacking is the right decision, that is, he assigns a positive probability to the alert not being triggered. If the alert is not triggered, he knows that no agent will attack afterwards and the regime will survive. Thus, there is always a strictly positive benefit from waiting for the alert. However, waiting is costly. If the principal sets the alert in a timely manner, that is, before some threshold time $\hat{\tau}$, then it follows from the continuity that the waiting cost will be arbitrarily small. Thus, under a *timely* disaster alert policy, the expected benefit always outweighs the expected cost of

forward induction refinement, we do not impose any constraint on the beliefs conditional on unexpected histories. We discuss this in section 5.1.

waiting.

When the principal sets a timely disaster alert, agents not only follow the principal's recommendation after the alert, but also always wait for the alert regardless of their signals. This implies that any regime that could have survived if no agent had attacked will indeed survive in the end. In other words, timely disaster alert eliminates panic.[4]

In a coordination problem, the agents face uncertainty about the fundamentals and other agents' strategies. However, if there is an early warning system in place that sends a warning before disaster occurs, then the agents will wait for the warning, and all unnecessary panic can be avoided. This shows that panic is a fragile idea. If the agents are sequentially rational and believe that other agents are sequentially rational as well, the principal can manipulate the agents and stop them from panicking in a simple manner. The principal does not need to know the private signals each agent receives. She uses a public and binary disclosure policy, and achieves her most preferred outcome. It is worth pointing out that the policy does not violate the principal's ex-post incentive compatibility. To observe this, note that the regime is doomed to fail when the alert is triggered regardless of the message the principal sends. This means that the principal does not need an ex-ante commitment to implement such a policy.[5]

After the financial crisis of 2008, bank supervisors began to adopt stress tests, partially as an information disclosure policy, in banking regulation.[6] In practice, a stress test is designed to evaluate whether a bank, or financial institute, is well equipped to deal with stress scenarios that may arrive in the future. We apply our main result to design a forward-looking stress tests that can eliminate bank run. To this end, in section 4, we modify the benchmark setup. Similar to our benchmark setup, nature selects the severity of the shock at the beginning of the game and agents receive noisy signals regarding the bank's preparedness to deal with the shock. However, unlike in our benchmark setup, the shock arrives at some point within the time window $[0, T]$ and not necessarily in the beginning.

Although the principal may not know when exactly the shock will arrive, she is able

[4]This insight can be extended to a case where agents receive more information from outside sources over time. The extended policy in such a situation is to set the timely disaster alert right after the arrival of any new information, which could potentially induce a panic (see section 3 for details).

[5]However, in practice, this may not always be the case. A principal may want to reduce the attack even when she knows that a regime switch is inevitable. Also, the principal may have a time preference — for example, she may want to defend the regime as long as possible. In these cases, implementing a disaster alert requires ex-ante commitment.

[6]*"The plan aimed to impose transparency on opaque financial institutions and their opaque assets in order to reduce the uncertainty that was driving the panic."*, Timothy Geithner, Stress Test: Reflections on Financial Crises

to do due diligence to determine how prepared the bank is when the shock arrives. This captures the forward-looking feature in bank stress tests. In this setup, a timely disaster alert is a quick stress test that sends a warning to the investors if the bank's balance sheet is so weak that *the bank cannot survive the shock when the shock arrives* regardless of the investors' actions in the future.

Note that, unlike the slow-moving capital outflow, investors can withdraw immediately in a financial market. Thus, in such a fast-moving market, a regime can change at any point in the middle of the time window. As soon as enough investors leave, the regime changes, and thus, an investor may incur substantial cost even if he waits for a little while. The effectiveness of such a stress test depends on when the shock arrives. If the shock arrives before the time of disclosure, it might be too late for investors to act on it. However, a timely stress test can limit this chance and thus, it is very likely to work as an early warning. This means that on expectation, the cost of waiting for the stress test is very small. Thus, all agents would wait for the test and then act according to the test result. This means that a forward-looking stress test completely eliminates panic based runs on banks when it is done in a timely manner.

Related Literature

Our study improves the understanding of how manipulating information can help in coordination. The studies most closely related to our work are by Goldstein and Huang (2016) and Inostroza and Pavan (2018), who also address the information design problem in a regime change game with privately informed agents. However, unlike our study, the authors consider a static game and the designer commits to an information disclosure rule about the underlying exogenous fundamental state.[7] While Goldstein and Huang (2016) proposes a simple stress test policy, Inostroza and Pavan (2018) designs the optimal disclosure policy, which, under some conditions, can be a stress test. However, in general, the principal may benefit from discriminatory disclosure.[8] In contrast, we consider a dynamic regime change game, in which the principal can also disclose information about the history of attacks, which is endogenous.

[7] Edmond (2013) also considers a static regime change game. However, unlike these studies, the principal can pay a cost to exaggerate the strength of the regime.

[8] Li, Song and Zhao (2019) considers a world where agents share a common belief. The authors construct an optimal information disclosure policy that partially discloses the true state to a fraction of agents and locally exaggerate the state to the rest of the agents.

The idea of the stress test in the aforementioned studies is slightly different from ours. It discloses whether the fundamental is beyond a certain threshold. This leads to multiple equilibria, and the studies focus on the worst equilibrium in which agents attack most aggressively. Goldstein and Huang (2016) and Inostroza and Pavan (2018) show that, when the stress test is tough enough — that is, the threshold is high enough — then, even in the worst equilibrium, the agents will not attack a regime that passes the stress test. However, if the bank fails the stress test, then investors will withdraw and the bank will fail. Note that the bank's fundamental may be strong enough such that a run is not warranted, but it may not be strong enough to pass a tough stress test. Thus, panic cannot be eliminated by disclosing information only about the exogenous state. In sharp contrast, we show that, if there is a time window, and the principal can disclose information about the endogenous history as well, then panic can be eliminated.[9]

This study contributes to the recent and growing literature on dynamic information design. This literature (See Au (2015), Orlov, Skrzypacz and Zryumov (2019), Ely and Szydlowski (2019), Ball (2019), Kolb and Madsen (2019)) primarily considers the persuasion of a single agent by disclosing information about an exogenously evolving state. The principal conditions the future disclosure on the current action of the agent. Che and Hörner (2018) considers multiple agents deciding whether to experiment with a new product but there is no payoff externality. The principal sends spam recommendation to agents arriving early to facilitate social learning. We consider a mass of agents who play a dynamic regime change game. We borrow this game from Morris and Shin (2003) and in tradition of Carlsson and Van Damme (1993), we consider the agents to be privately informed. However, we allow for a general information structure. Similar to Dasgupta (2007) and Dasgupta, Steiner and Stewart (2012), we extend the canonical regime change game to allow for an endogenous delay in attack. Unlike the aforementioned dynamic information design studies, the principal manipulates the agents by promising to disclose information in the future regarding the action by other agents.

The remaining of the paper is organized as follows. Section 1 describes the model. Section 2 demonstrates that a timely disaster alert eliminates panic. Section 3 shows that the result can be extended to the arrival of new information over time. Section 4 applies the theory to design a forward-looking stress test. Section 5 discusses which features of the

[9]In a related work, Basak and Zhou (2020) considers the case when agents move sequentially in an exogenous order. This exogenous timing specification was motivated by the observation that, borrowers often adopt a staggered debt structure in which the short-term debts mature at different dates. The crucial option value argument that we use in this study does not apply when timing is exogenous.

model are essential and which ones could be relaxed and section 6 concludes. Some of the proofs have been relegated to the appendix.

1 Model

Players and Actions The economy is populated by a principal, a continuum of agents, indexed by $i \in [0,1]$, and a regime. A shock hits the regime and it is commonly known. We normalize the time at which the shock arrives as 0. Once the shock arrives, the agents are allowed a time window $[0,T]$ to decide whether they want to attack the regime. [10]

Let us denote the action of attacking by 1 and not attacking by 0. An agent i chooses $a_i \in [0,T]^{\{0,1\}}$ which describes whether he attacks at any time. Attacking is an irreversible action, whereas not attacking is reversible. Hence, if agent i has already attacked by some t, he has no more decisions to make. However, if he has not attacked, then he has the option to attack at any time between t and T, or not attack at all. If agent i decides to attack at time $t_0 \in [0,T]$, then $a_{it} = 0$ for any $t \in [0,t_0)$ and $a_{it} = 1$ for any $t \in [t_0,T]$. For any agent i who decides to attack, that is, $a_{iT} = 1$, let us denote the time of attack as $t_i \equiv \min\{t \in [0,T] | a_{it} = 1\}$. The time of attack is defined to be $t_i = \infty$ for an agent i who does not attack at all. Hence, a_i can simply be represented by the time of attack $t_i \in [0,T] \cup \infty$. At any t within the time window, the mass of agents who already attack is

$$N_t \equiv \int_{i\in[0,1]} \mathbf{1}\{i | t_i \leq t\} di.$$

By definition, the mass of attacks $N_t \in [0,1]$ is (weakly) increasing in time t.

Fundamental State The underlying state of the economy is captured by θ. We refer to it as the fundamental strength of the regime. It captures the preparedness of the regime to face the shock. At time 0, nature draws a state $\theta \in \Theta$, where Θ is a compact subset of $\mathbb{R}$. It is common knowledge that θ is drawn from some distribution Π with smooth density π strictly positive over Θ.

Regime Outcome Let $r \in \{0,1\}$ denote the fate of the regime. We represent the event that the regime survives by $r = 0$ and the complementary event by $r = 1$. The fate of

[10] In different applications of a regime change game, attacking could mean exiting from a market, withdrawal of early investment, shorting a currency, or making redemption from a mutual fund.

the regime is decided at T depending on the fundamental state (θ) and the aggregate attack until the end (N_T). The regime survives — that is, $r = 0$ — if and only if $R(\theta, N_T) \geq 0$, where $R(.)$ is a continuous function that is increasing in θ and decreasing in N_T.[11]

Payoff The agents are ex-ante identical and expected utility maximizers. If an agent does not attack ($t_i = \infty$), then he obtains

$$v(\theta, N_T) = \begin{cases} g(\theta, N_T) & \text{if } r = 0 \\ l(\theta, N_T) & \text{if } r = 1, \end{cases}$$

and if he attacks at time t, then he obtains $u(t)$. We normalize $u(0) = 1$. As is standard in the static regime change game (e.g., Inostroza and Pavan (2018)),

$$g(\theta, N_T) > 1 > l(\theta, N_T).$$

This captures how, if the regime is going to survive ($r = 0$), then not attacking is the desirable action; and if the regime is not going to survive ($r = 1$), then attacking is the desirable action. We allow $g(.)$ and $l(.)$ to be non-monotonic in the arguments, but there exists $\underline{g} > 1$ such that $g(.) \geq \underline{g}$ and there exists $\underline{l} < \bar{l} < 1$ such that $\underline{l} \leq l(.) \leq \bar{l}$.

Different from the static payoff case, the agent has multiple opportunities to attack and delaying the attack is costly, that is, $u(t)$ is decreasing in t.[12] We assume that $u(T) > \bar{l}$. This means that, even at the last minute, if the agents learn that the regime will not survive ($r = 1$), attacking is the desirable action. We further assume that $u(t)$ is *Lipschitz continuous* in t.

Dominance Region There exists $\underline{\theta}, \bar{\theta} \in \Theta$ such that $R(\underline{\theta}, 0) = R(\bar{\theta}, 1) = 0$. This means that, when $\theta \in \Theta^L = \Theta \cap (-\infty, \underline{\theta})$, the regime cannot survive regardless of whatever strategy the agents adopt; and when $\theta \in \Theta^U = \Theta \cap [\bar{\theta}, +\infty)$, the regime will always survive regardless of whatever strategy the agents adopt. We refer to Θ^U (or Θ^L) as the upper (or lower) dominance region, where not attacking $t_i = \infty$ (or attacking right away $t_i = 0$) is

[11] In section 4, we relax this assumption when we consider a financial market. In a financial market, withdrawals can be instantaneous. Hence, the regime can change at any time t in the middle of the time window as soon as $R(\theta, N_t) < 0$.

[12] In the online appendix, we construct a simple example of capital outflow to study the case with flow payoff. In this example, waiting is not necessarily costly. However, a capital tax policy can make waiting costly and complement the information disclosure policy.

the dominant strategy. We assume that $\Theta^U \neq \emptyset$.

Exogenous Information In addition to the common prior Π, each agent i receives a signal $s_i \in \mathbb{S}$ before he decides when to attack (if at all). Given any underlying fundamental θ, the signal profile $s(\theta) \in \mathbb{S}^{[0,1]}$ is drawn from a distribution $F(s|\theta)$ with associated density $f(s|\theta)$. Note that this allows for any arbitrarily correlated signal, ranging from conditionally independent private signals to public signals. For simplicity, we assume that the agents do not receive any more information over time about the fundamental or observe other agents' actions. In section 3, we relax this assumption and allow agents to receive additional information.

The following assumption restricts the information-generating process F. It holds that, regardless of whatever noisy signal an agent receives, he always assigns some positive chance that $\theta \in \Theta^U$. In other words, regardless of his signal and other agents' actions, he always doubts that if attacking is the right decision.

Assumption 1 *(Doubt) There exists $\epsilon > 0$ such that, any agent i with noisy signal s_i believes that*

$$\mathbb{P}(\theta \in \Theta^U|s_i) = \frac{\int_{\theta\in\Theta^U} f_i(s_i|\theta)\pi(\theta)d\theta}{\int_{\theta\in\Theta} f_i(s_i|\theta)\pi(\theta)d\theta} > \epsilon,$$

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

In particular, if f_i has full support, and it is bounded away from 0, then the above assumption holds true. [13]

Principal The principal's payoff depends only on whether the regime survives: 1 if the regime survives and 0 if it does not.[14] The principal knows θ and also observes N_t at any time t. However, she does not have access to the agents' noisy private information.

Disclosure Policy The principal commits to a dynamic information disclosure policy. For any $\tau \in [0, T]$, the principal can disclose some information to the agents based on the exogenous fundamental θ and the endogenous attack so far N_τ. We consider a continuous time model, that is, an agent who has not attacked by time τ, has an opportunity to attack

[13] A sufficient condition that validates this doubt assumption is that $f_i(s_i|\theta) > \frac{\epsilon}{1-\Pi(\bar{\theta})}$ for all $i \in [0,1]$, $s_i \in \mathbb{S}$ and $\theta \in \Theta^U$.

[14] It is easy to generalize to the case where the principal also wants to minimize the aggregate attack conditional on the regime's survival.

again at $(\tau + dt)$, where $dt \to 0$. At any time τ we allow for a sequence of events to occur. Accordingly, we define τ^- and τ^+. First, at τ^-, an agent can attack, then at τ^+, the principal can disclose information. There is no time discounting between τ^- and τ^+. We use the convention that even time zero disclosure is after the agents move at time zero. This means that even time zero disclosure can contain information regarding the history of play.[15]

Let $\mathcal{S}$ be a compact metric space defining the set of possible disclosures to the agents, and $m_i(\tau, \theta, N_\tau) \in \mathcal{S}$ be the message to agent i. A general disclosure policy $\Gamma = (q, \mathcal{S})$ consists of a set of disclosed messages $\mathcal{S}$ and the disclosure rule $q : [0, T] \times \Theta \times [0, 1] \to \Delta(\mathcal{S}^{[0,1]})$. The feature of endogenous move enables the principal to select the time of disclosure, and to make this information disclosure policy history dependent.

Robust Design We use extensive form rationalizability à la Pearce (1984) as our solution concept.[16] Given the disclosure policy Γ, let $\mathcal{R}(\Gamma)$ be the set of all possible rationalizable strategy profiles $a \equiv (a_i(s_i))$. Define

$$\Theta^F(\Gamma) := \{\theta \in \Theta | R(\theta, N_T(a)) < 0 \text{ for some } \ a \in \mathcal{R}(\Gamma)\}.$$

Thus, if $\theta \notin \Theta^F(\Gamma)$, then the regime will survive regardless of whatever rationalizable strategies the agents play; if $\theta \in \Theta^F(\Gamma)$, then the regime may not survive. The principal's objective is

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

where $\mathbb{P}(\theta \in \theta^F(\Gamma)) = \int_{\Theta^F(\Gamma)} d\Pi(\theta)$. That is, the principal anticipates, state by state, the "worst possible" outcome that is consistent with some rationalizable strategy profile, and chooses the policy Γ to minimize the ex-ante chance that the regime may not survive.

Note that, when $\theta \in \Theta^L$, the regime fails irrespective of the size of the attack. Hence, any disclosure policy Γ cannot endure such a regime, that is, $\Theta^L \subseteq \Theta^F(\Gamma)$. A regime could also fail even when it is not warranted ($\theta \notin \Theta^L$), because agents attack thinking that others will attack. We refer to this as *panic-based attacks*. Let us define $\Theta^P(\Gamma) := \Theta^F(\Gamma) \backslash \Theta^L$ for any policy Γ as the set of fundamental in which the regime can fail because of panic-based

[15] One can also allow the principal to disclose information before the game begins. But in this case, the disclosure will be only about the exogenous state.

[16] We only assume that the agents commonly believe in sequential rationality in the beginning. We do not impose any restriction on beliefs conditional on unexpected histories. We discuss this in detail in section 5.1.

attacks. If $\Theta^P(\Gamma) = \emptyset$, then we can claim that the policy Γ eliminates panic.[17]

2 Main Result

We restrict our attention to a simple dynamic information disclosure policy. This policy induces the agents to perfectly coordinate their actions and never attack a regime when it is not warranted ($\theta \notin \Theta^L$), and thus, eliminates panic.

Disaster Alert

We refer to the following disclosure policy as *disaster alert*. The principal only discloses information once at some time $\tau \in [0, T]$. The public signal d^τ is generated based on the underlying fundamental θ and the history of attacks N_τ as follows:

$$d^\tau(\theta, N_\tau) = \begin{cases} 1 \text{ if } R(\theta, N_\tau) < 0 \\ 0 \text{ otherwise.} \end{cases}$$

We denote this binary public disclosure policy as Γ^τ. Upon receiving the signal $d^\tau = 1$, the agents understand that the regime cannot survive in the end, that is, $R(\theta, N_T) < 0$ (since N_t is weakly increasing in t). Hence, for agents who have not attacked, it is the dominant strategy to attack at time $\tau + dt$. In this sense, $d^\tau = 1$ acts as an alert for disaster. On the other hand, if the alert is not triggered, or $d^\tau = 0$, the agents understand that $R(\theta, N_\tau) \geq 0$, and thus, the regime will survive if no agent attacks the regime after time τ.[18]

Option Value of Waiting

Under the policy Γ^τ, the agents will only have one chance to receive new information at τ^+. Hence, attacking at any time $t_i \in (0, \tau^-]$ is dominated by attacking at time $t_i = 0$, since waiting is costly. Similarly, after receiving the new information d^τ, attacking at anytime $t_i \in (\tau + dt, T]$ is dominated by attacking immediately after the disclosure at $\tau + dt$.

[17] Doval and Ely (2019) extends the Bayes correlated equilibrium idea of Bergemann and Morris (2016) for extensive form game and call it the coordinated equilibrium. It is correlated equilibrium in which all the agents perfectly coordinate on not attacking when a run is not warranted. Note that if $\Theta^P(\Gamma) = \emptyset$, then under policy Γ, it is not only a possible equilibrium, but it is the only equilibrium.

[18] In the spirit of Bayesian persuasion, this can be thought of as the principal sending a recommendation at time τ to the agents to "attack" when the disaster alert is triggered, and "not attack" otherwise.

Lemma 1 *(Option Value) Under the disclosure policy Γ^τ, for any noisy signal s_i, the only rationalizable strategies are*

$\mathcal{A}$: *attack at time* 0*, that is,* $a_{it}(s_i) = 1$ *for all* $t \in [0, T]$*, and*

$\mathcal{W}$: *wait until time* τ *for the disaster alert, that is,* $a_{it}(s_i) = 0$ *for all* $t \in [0, \tau + dt)$*, and then follow the alert, that is,* $a_{it}(s_i, d^\tau) = d^\tau$ *for* $t \in [\tau + dt, T]$*.*

Proof. First, it is not rational for agents to attack at any time other than 0 and $\tau + dt$. Second, when $d^\tau = 1$, the dominant action is attack. Hence, the only possible strategies are: $\mathcal{A}$, $\mathcal{W}$, and attacking at time $\tau + dt$ independent of d^τ. Let us call this third strategy $\mathcal{W}'$. Note that the strategy of $\mathcal{W}'$ generates a payoff of $u(\tau + dt)$, which is strictly less than $u(0)$. Thus, it is strictly dominated by $\mathcal{A}$. ■

For any agent who decides to wait for the disclosure (instead of attacking immediately), the information that will be disclosed at time τ must be valuable to him. This means that he will never take the same action (attack) regardless of the information received in future. Otherwise, there is no option value associated with the information arriving in the future, and hence he will not wait. Chamley and Gale (1994) and Gul and Lundholm (1995) make a similar argument in the context of social learning in which an agent can learn from others' actions, but such actions do not affect his payoff. [19]

In this study, the principal controls the information flow after time 0. If the disaster alert is triggered ($d^\tau = 1$), then attacking is the dominant strategy for an agent. Therefore, for the positive option value of waiting, it must hold that the agent will not attack when the alert is not triggered.

This implies that when the disaster alert is not triggered ($d^\tau = 0$), the regime will survive in the end. This is because there is no further attack after time τ when $d^\tau = 0$, or $N_T = N_\tau$. Thus, no alert ($R(\theta, N_\tau) \geq 0$) implies the survival of the regime, that is, $R(\theta, N_T) \geq 0$. Consequently, under the policy Γ^τ, the agents who decide to wait for the information disclosure perfectly coordinate their actions, and the strategic uncertainty is completely removed from time τ onward.

[19] The intuition is simple: Consider two agents deciding whether to attack at time 1 or time 2. If an agent waits 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, there is no positive option value of waiting.

Timely Disaster Alert

Lemma 2 *(Timely Alert) There exists $\hat{\tau} > 0$ such that, under the disclosure policy Γ^τ, where $\tau < \hat{\tau}$, for any signal structure satisfying Assumption 1, the only rationalizable strategy for an agent with signal s_i is $\mathcal{W}$ — that is, wait for the disaster alert, and then follow the alert.*

Proof. Consider an agent who has decided to wait for the disclosure (plays $\mathcal{W}$). If $R(\theta, N_\tau) < 0$, the alert triggers ($d^\tau = 1$). Then, he attacks at time $\tau + dt$ and receives $u(\tau + dt)$. Otherwise, $R(\theta, N_\tau) \geq 0$, and the alert does not trigger ($d^\tau = 0$). We know from Lemma 1 that the agent will not attack. Thus, the expected payoff from playing $\mathcal{W}$ is

$$\mathbb{P}(d^\tau = 1|s_i)u(\tau + dt) + \mathbb{P}(d^\tau = 0|s_i)E\left(v(\theta, N_T)|s_i, d^\tau = 0\right).$$

If $d^\tau = 0$, it follows from Lemma 1 that no agent who has waited will attack, or $N_T = N_\tau$. This implies that, when the alert is not triggered ($d^\tau = 0$), $R(\theta, N_T = N_\tau) \geq 0$, that is, the regime survives ($r = 0$). Recall that, if $r = 0$, then by not attacking, the agent receives $g(\theta, N_\tau)$. Finally, since $dt \to 0$, the expected payoff from playing $\mathcal{W}$ simplifies to

$$\mathbb{P}(d^\tau = 1|s_i)u(\tau) + \mathbb{P}(d^\tau = 0|s_i)E\left(g(\theta, N_\tau)|s_i\right).$$

On the other hand, the expected payoff from attacking immediately ($\mathcal{A}$) is $u(0) = 1$. Hence, the expected payoff difference from strategy $\mathcal{W}$ compared with $\mathcal{A}$ is

$$D(\Gamma^\tau, s_i) = \mathbb{P}(d^\tau = 1|s_i)(u(\tau) - u(0)) + \mathbb{P}(d^\tau = 0|s_i)(E(g(\theta, N_\tau)|s_i) - 1). \tag{1}$$

Since $u(t)$ is Lipschitz continuous, $u(0) - u(\tau) \leq \mathcal{C}\tau$ for some positive finite $\mathcal{C}$. It follows from Assumption 1 that $\mathbb{P}(d^\tau = 0|s_i) > \epsilon$. Also, recall that $g(\theta, N) \geq \underline{g} > 1$. Therefore,

$$D(\Gamma^\tau, s_i) > -(1 - \epsilon)\mathcal{C}\tau + \epsilon(\underline{g} - 1).$$

Define

$$\hat{\tau} := \frac{\epsilon}{1 - \epsilon}\left(\frac{\underline{g} - 1}{\mathcal{C}}\right).$$

If $\tau < \hat{\tau}$, then $D(\Gamma^\tau, s_i) > 0$ for any s_i. This implies that an agent prefers to wait for the alert and then follows the principal's recommendation ($\mathcal{W}$) rather than attacking right away ($\mathcal{A}$), regardless of his private signal and other agents' actions. ■

The tradeoff that agents face when choosing between attacking immediately ($\mathcal{A}$) and the strategy of "wait and see" ($\mathcal{W}$) is as follows: A cost $(u(0) - u(\tau))$ is associated with a delayed attack when the alert ($d^\tau = 1$) is triggered. If the alert is not triggered ($d^\tau = 0$), the regime survives in the end regardless of other agents' actions (either $\mathcal{A}$ or $\mathcal{W}$). Hence, adopting the strategy of "wait and see" ($\mathcal{W}$) can prevent agents from making a mistake by moving early, that is, attacking a regime that survives. The benefit from this more informed choice is at least $(\underline{g} - 1)$. Assumption 1 guarantees that regardless of the signal s_i and other agents' actions, an agent assigns positive probability to the fact that the alert will not be triggered. In that case, attacking early is definitely a mistake. This means that there is a strictly positive benefit of waiting.

Lemma 2 shows that, if the disaster alert can be set in a timely manner, the cost of the delay will be limited (strictly lower than the expected benefit), which makes the strategy of "wait and see" ($\mathcal{W}$) a strict dominant strategy.

We use the solution concept of extensive form rationalizability. However, the argument only relies on the assumption that an agent is sequentially rational, and he believes that others are sequentially rational. It does not require higher order beliefs, such as the agent's belief that other agents believe that others are sequentially rational. As long as an agent is sequentially rational, the option value argument holds (Lemma 1), that is, an agent will take either $\mathcal{A}$ or $\mathcal{W}$. If he believes that other agents are sequentially rational, he understands that the regime will not fail if the disaster alert is not triggered.[20] For a timely disaster alert, under the assumption 1, the expected payoff from $\mathcal{W}$ is strictly higher than $\mathcal{A}$ regardless of whether others play $\mathcal{A}$ or $\mathcal{W}$. Thus, if an agent believes that others are rational, he will take the strategy $\mathcal{W}$ rather than $\mathcal{A}$.

No Panic

Theorem 1 *(No Panic) Under the disclosure policy Γ^τ with $\tau < \hat{\tau}$, if the information structure satisfies Assumption 1, then there is no panic, that is,*

$$\Theta^P(\Gamma^\tau) = \emptyset.$$

Proof. The only rationalizable strategy is $\mathcal{W}$ when $\tau < \hat{\tau}$ (lemma 2). Thus, $N_\tau = 0$. For any regime with $\theta \notin \Theta^L$, $R(\theta, N_\tau = 0) \geq 0$. Hence, no alert will be triggered and no

[20] If he does not believe that all others are sequentially rational (or only playing $\mathcal{A}$ and $\mathcal{W}$), then this argument may not hold. We discuss this issue in section 5.

further attack occurs, or $N_T = N_\tau = 0$ (Lemma 1). Therefore, any regime with $\theta \notin \Theta^L$ survives (since $R(\theta, N_T) \geq 0$). Hence, $\Theta^F(\Gamma^\tau) = \Theta^L$ and $\Theta^P(\Gamma^\tau) = \emptyset$. ■

Theorem 1 follows immediately from Lemma 1 and Lemma 2. Under a timely disaster alert, all agents would wait for the disclosure and follow the alert thereafter regardless of their noisy signals. Since all agents are waiting, any regime that can survive without attack ($\theta \notin \Theta^L$) will not trigger the alert. Since the alert is not triggered, the agents will follow the alert and not attack. Thus, any regime that can survive without any attack will survive for sure. In other words, a timely disaster alert eliminates panic. Note that this simple public disclosure policy makes the principal's favorite outcome the unique rationalizable outcome.

Under strategic complementarity, "runs" are common. Runs could be based on the fundamental ($\theta \in \Theta^L$), but they could also arise from panic. The principal cannot save a regime when attacks are purely based on the fundamental, but the above theorem shows that panic can be eliminated. More importantly, panic can be eliminated using a very "simple" disclosure policy. Recall that the principal does not have access to the agents' private signals. The disaster alert policy does not require disclosure conditional on such private signals. Moreover, the policy is a binary and public disclosure — whether the alert is triggered — and it does not send private messages to the agents. Finally, the alert is triggered only when $R(\theta, N_\tau) < 0$, that is, the regime cannot survive regardless of the agents' actions. This means the principal cannot receive a higher payoff by misreporting. Thus, the principal's ex-post incentive compatibility holds.

History Dependent Disclosure

It is important to note that the disaster alert is a partial disclosure policy that depends on the history. The literature focuses on the static regime change game and the disclosure is only about the exogenous fundamental θ.

Consider an alternative disclosure policy $\tilde{d}$, which publicly discloses whether $\theta \geq \underline{\theta}$ before the game begins. Then, one possible equilibrium outcome is that no agent attacks, and the regime survives. However, this is not the unique rationalizable strategy. If an agent receives a low signal and believes others will attack, then he will attack as well. In fact, Angeletos, Hellwig and Pavan (2007) shows that there are many other possible equilibria in which the regime could fail because of panic-based attacks. Goldstein and Huang (2016) and Inostroza and Pavan (2018) argue that this disclosure policy is not strong enough to dissuade the agents from attacking in the worst possible equilibrium. The authors propose a stress test policy — disclose whether $\theta \geq k$, where $k \geq \underline{\theta}$. They show that k is large enough, then even in the worst possible equilibrium, the agents will not attack the regime that passes the test ($\theta \geq k$). Note that if $\theta \in [\underline{\theta}, k]$, then the regime is not so weak that a run is not warranted, however, it is not so strong either that it can pass the tough stress test. These regimes will fail because of panic based runs. Thus, a history independent disclosure before the game begins cannot eliminate panic.

Now suppose that the principal discloses the same information as in $\tilde{d}$, but at time τ rather than before the game begins. Let us call this new policy $\tilde{d}^{\tau}$. Note that unlike our proposed disaster alert this is a history-independent disclosure. As in Lemma 1, an agent who waits for the disclosure will not attack after learning that $\theta \geq \underline{\theta}$. However, unlike the case of the disaster alert, if the agents attack before the alert, then the regime may not survive even after it is disclosed that $\theta \geq \underline{\theta}$ at time τ. This means that the benefit from waiting for the alert may not be strictly positive. The following simple example illustrates this.

Example 1 *Agents have a homogeneous belief that* $\mathbb{P}(\theta \geq \bar{\theta}) = \mathbb{P}(\theta < \underline{\theta}) = \epsilon > 0$ *and* $\mathbb{P}\left(\theta \in (\underline{\theta}, \bar{\theta})\right) = 1 - 2\epsilon$. *The loss and gain are constant:* $l(\theta, N) = l$ *and* $g(\theta, N) = g$. *If all agents attack at time* 0*, the expected net payoff from waiting and only attacking when* $\theta < \underline{\theta}$ *compared with attacking at* 0 *is*

$$\epsilon(u(\tau) - u(0)) + (1 - 2\epsilon)(l - 1) + \epsilon(g - 1).$$

If $\epsilon < \frac{1-l}{g-1+2(1-l)}$*, then this net benefit is strictly negative regardless of the time of disclosure* τ*. Thus, if* ϵ *is sufficiently small, then under the history-independent disclosure policy* $\tilde{d}^{\tau}$*, a possible equilibrium is one in which all the agents attack at time* 0*.*

Note that, in the above example, the doubt assumption is satisfied, and therefore, a timely disaster alert policy d^{τ} eliminates panic. However, the history-independent disclosure policy $\tilde{d}^{\tau}$ cannot eliminate panic regardless of τ. The difference is the second term in the above net benefit expression $(1-2\epsilon)(l-1)$. When $\theta \in [\underline{\theta}, \bar{\theta})$ and all the agents attack at time 0, the regime cannot survive. Under the disclosure policy d^{τ}, agents receive the alert that the regime is doomed to fail. However, under the disclosure policy $\tilde{d}^{\tau}$, agents only learn that the regime could have survived had agents refrained from attacking.

3 New Information and Repeated Disaster Alert

In our benchmark setup, we assume that the agents do not receive 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 from other sources. In this case, agents cannot be certain that, even if the timely disaster alert is not triggered, the agents will not panic later and attack when they receive new information. This could make the disaster alerts ineffective. Here, we consider the optimal disclosure policy under the exogenous arrival of new information.

As before, we assume that at the beginning of the game, nature draws the state θ. However, unlike the benchmark setup, the agents get noisy signals not only in the beginning of the game, but also at different dates over the time window $[0, T]$. We assume that information does not arrive very frequently. In fact, there are only a finite number of dates at which new signals can arrive in the time interval $[0, T]$. Let us denote these dates as $\{t_0, t_1, \ldots t_K\}$, where $t_0 = 0$. Define $\overline{\tau} := \min_{j=0}^{K+1}\{t_j - t_{j-1}\}$, where $t_{K+1} = T$. Thus, whenever the agents receive some information, there is at least a time window of $\overline{\tau}$ before a new information arrives or the game ends.

We use the convention that the news arrives at t_k^-, and the agent who receives a signal s_i^k at time t_k^- can act based on this news as early as at time t_k, $k = 0, 1, \ldots K$. The profile of signals $s^k(\theta)$ is drawn according to some distribution $F^k(s^k|\theta)$ with associated density $f^k(s^k|\theta)$.

Suppose that (some) agents receive multiple signals from outside sources at t_k, $k = 0, 1, \ldots K$. As before, we assume that the agents have doubt, that is,

$$\exists \epsilon > 0 : \mathbb{P}(\theta \in \Theta^U | s_i^0, s_i^1, s_i^2, \ldots, s_i^K) > \epsilon \text{ for any } s_i^0,\ \ s_i^1, s_i^2, \ldots, s_i^K.$$

Since the exogenous information does not arrive very frequently, it gives the principal the scope to set a timely disaster alert each time after agents receive new information. With the new arrival of exogenous information, a natural extension of the one-shot disclosure policy is to set the disaster alert right after any new information arrives. To reduce the burden of notation, we use the same notation Γ^τ to capture this modified policy – there are $(K + 1)$ disaster alerts at time $(t_k + \tau)^+$ for $k = 0, 1, \ldots, K$.

Theorem 2 *When new information arrives over time, under the extended disclosure policy Γ^τ with $\tau < \min\{\hat{\tau}, \bar{\tau}\}$, there is no panic.*

To understand the argument consider the following simple case: $K = 1$. Suppose agent i receives a noisy private signal s_i^0 at time 0^- (as before) and s_i^1 at time t_1^- for some $t_1 \in (0, T]$ about the fundamental θ. The agents can act based on the private signal they receive as early as in the same period, that is, at time 0 and t_1. The basic argument is an extension of our main result. First, if the first alert was triggered at time τ, that is, $d^\tau = 1$, then all the agents attack and there is no more decision to be made after the new information arrives. Let us assume otherwise and begin our analysis from time t_1. Consider any agent who has not attacked before time t_1 and receives new signal s_i^1. Based on the same argument as in Lemma 1, for any possible signal s_i^0 and s_i^1, agent i either attacks at time t_1, or waits for the second alert, and then attacks if and only if $d^{t_1+\tau} = 1$.

Under the extended doubt assumption, agents cannot be sure that the regime will fail if all other agents attack after receiving additional information. Therefore, we can apply the same argument as in Lemma 2 to show that, with a timely disaster alert ($\tau < \min\{\hat{\tau}, \overline{\tau}\}$), any agent who does not attack (after seeing no alert being triggered at time τ, or $d^\tau = 0$) will wait for the next disclosure ($d^{t_1+\tau}$) and only attack when the second alert is triggered ($d^{t_1+\tau} = 1$).

Hence, if the first alert is not triggered ($d^\tau = 0$), any agent who has waited for the first alert will not attack after receiving the new information at t_1, but would wait for the second alert and follow it. In this case, $d^{t_1+\tau} = d^\tau = 0$. Otherwise, if the first alert is triggered, all agents who have not attacked would attack immediately, and thus, the second alert becomes irrelevant. Hence, given that the timely disaster alert in the future can always prevent agents from acting on any new information, the problem faced by agents is essentially the same as the one in the benchmark setup. Agents will never attack based on their private information, rather wait and following the alerts. Thus, as long as any additional learning cannot remove the doubt, the extended timely disaster alert policy eliminates panic.

4 Liquidity Crisis and Forward-looking Stress Test

The bank run is an prevalent examples of the regime change game. There are two important features of bank run which are missing in our benchmark setup. First, unlike in our benchmark setup, a regulator is capable of acquiring information to learn the severity of the macro shock that would arrive in the future, and she can examine the bank's balance sheet to determine θ, the preparedness of the bank to deal with the macro shock when it arrives. Thus, the principal can disclose information based on θ, even before the shock arrives. We call this modification *Due Diligence*.[21]

Second, in the benchmark model, we assume that the regime change occurs in the end of a time window. This is a reasonable assumption for slow-moving capital market. However, in a financial market, withdrawal of funds needs to be executed immediately. At any point in time, if enough investors have already made withdrawals, then the financial institution may fall into a liquidity crisis and fail to service withdrawal requests further. This means that any delay in making withdrawals (or attacks) may cause a significant loss if many other investors withdraw during this period of delay, thus rapidly exhausting the financial institution's liquidity and leading to a liquidity crisis. Therefore, the cost of waiting is not continuous in the financial market as we have assumed in our benchmark setup. This cost not only depends on the time of the withdrawal request, but also on whether the liquidity crisis kicks in during the period of delay because of other agents' withdrawal decisions. We refer to this modification as *Sequential Service Constraint*.

Stochastic arrival of shock At the beginning of the game, nature chooses θ, the preparedness of the bank to deal with a macro shock, and the agents receive some noisy signals. However, unlike the benchmark setup, the macro shock does not necessarily arrive at time 0. There is uncertainty regarding when this shock will arrive. We assume that the shock arrives at time t_s stochastically, following a commonly known distribution $G(t_s)$, with atom-less density, where $G(0) = 0$ and $G(T) = 1$.

Let θ_t capture the bank's fundamental at time t. We assume that unless the shock actually hits the bank, the bank's fundamental is strong enough to withstand any attack, that is, $\theta_t = \theta_0 > \overline{\theta}$ for $t \in [0, t_s)$. After the shock arrives, the fundamental strength is the bank's preparedness to deal with the shock, that is, $\theta_t = \theta$ for $t \in [t_s, T]$.

Sequential Service Constraint Once the shock arrives, the regime (or bank) can fail at any time $t \in [t_s, T]$ as soon as $R(\theta, N_t) < 0$. This can occur if the fundamental (or bank's liquidity position) θ is very low and/or many agents choose to attack (or withdraw) before time t. If the regime has changed by time t, then $r(t) = 1$; otherwise $r(t) = 0$. We assume that the investor can observe $r(t)$ (or the bank's failure) instantaneously at any t. We use the convention that, at any time t, first $r(t)$ is observed, and then the agents take their decisions.

[21] Orlov, Zryumov and Skrzypacz (2018) also assumes that, in bank stress tests, the regulator has an informational advantage by being able to obtain a forward-looking estimate of the systemic risk factor.

As before, if agents do not withdraw, the payoff $v(.)$ depends on the fate of the bank at the end of time window T.[22]

$$v(\theta, N_T) = \begin{cases} g \text{ if } r(T) = 0 \\ l \text{ if } r(T) = 1. \end{cases}$$

For simplicity, we assume that g and l are constant. The crucial difference from our benchmark setup is that, if an agent withdraws at time t, his payoff $u(.)$ depends on $r(t)$, that is, whether the bank has failed by time t.

$$u(t, \theta, N_t) = \begin{cases} u(t) \text{ if } r(t) = 0 \\ l \text{ if } r(t) = 1. \end{cases}$$

$u(t)$ is the payoff from withdrawing but only if the withdrawal occurs before the bank fails. As before, we assume that $u(t)$ is Lipschitz continuous, decreasing in t, $l < u(t) < g$, and we normalize $u(0) = 1$. After the bank fails, an agent receives l regardless of whether he withdraws or not. [23]

Therefore, if at any time t, an agent waits for next dt time and the regime changes during this waiting, then he receives l rather than $u(t + dt)$. The discontinuity in waiting cost is captured by the fact that $l < u(t + dt)$.

Forward-looking Stress Test Consider a disaster alert Γ^τ. Note that at time τ, it is possible that the shock has not arrived yet. Therefore, this disclosure requires due diligence. If $R(\theta, N_\tau) < 0$, then the bank will surely fail when the shock arrives (if it has not arrived already), and the disaster alert is triggered ($d^\tau = 1$). On the other hand, if $R(\theta, N_\tau) \geq 0$, then the bank will survive when the shock arrives if the agents who have not withdrawn already do not withdraw. In this banking context, we call the disaster alert policy a *forward-looking stress test*. If the alert is triggered, the bank has failed the test; otherwse, the bank has passed the test. Note that this is the weakest possible stress test in the sense that, if the bank fails this test, then it is doomed to fail whenever the shock arrives.

[22] One can interpret the liquidity problem as a temporary concern. After some date T, this concern is eliminated, as the bank can transform its illiquid assets into liquid ones or find extra funding sources to service all potential withdrawal requests.

[23] Another interpretation is that after the bank fails, all creditors will go through the bankruptcy process and they simply do not have the option to request withdrawals.

Theorem 3 *When the shock hits the bank stochastically following $G(.)$, there exists $\tilde{\tau}$ such that a forward-looking stress test Γ^τ with $\tau < \tilde{\tau}$ eliminates panic.*

Proof. As in Lemma 1, the option value argument holds, that is, an agent who has waited for the alert will attack if and only if the alert is triggered. Agents either take $\mathcal{A}$ or $\mathcal{W}$. If an agent waits for the alert, then the shock arrives before the scheduled alert with probability $G(\tau)$; with complementary probability, it will arrive later.

If the alert is not triggered, that is, $d^\tau = 0$, since all agents who have waited will not attack, the regime survives in the end, that is, $R(\theta, N_T) = R(\theta, N_\tau) \geq 0$ (Lemma 2)). If the alert is triggered and the shock has already arrived, then the investor will receive l if he takes $\mathcal{W}$. When the alert is triggered but the shock has not arrived yet (or the regime has not failed yet), agents who take $\mathcal{W}$ get $u(\tau + dt)$.

Therefore, the expected payoff from playing $\mathcal{W}$ is

$$\begin{aligned}
&G(\tau)\left(\mathbb{P}(d^\tau = 0|s_i)g + \mathbb{P}(d^\tau = 1|s_i)l\right)\\
+&(1 - G(\tau))\left(\mathbb{P}(d^\tau = 0|s_i)g + \mathbb{P}(d^\tau = 1|s_i)u(\tau + dt)\right)\\
\geq& G(\tau)l + (1 - G(\tau))\left(\epsilon g + (1 - \epsilon)u(\tau + dt)\right)
\end{aligned}$$

The inequality follows since $g \geq u(\tau + dt) \geq l$ and $\mathbb{P}(d^\tau = 0|s_i) > \epsilon$ (Doubt). On the other hand, the payoff from playing $\mathcal{A}$ is $u(0) = 1$. As in theorem 1, using the Lipschitz continuity of $u(.)$, we can say the expected net payoff from playing $\mathcal{W}$ compared with $\mathcal{A}$ is at least

$$G(\tau)(l - 1) + (1 - G(\tau))\left(\epsilon(g - 1) - (1 - \epsilon)\mathcal{C}\tau\right)$$

Hence, $\mathcal{W}$ dominates $\mathcal{A}$ if

$$\mathcal{C}\tau + \frac{G(\tau)}{1 - G(\tau)}\frac{1 - l}{1 - \epsilon} < \frac{\epsilon}{1 - \epsilon}(g - 1).$$

Since $G(0) = 0$ and $G(.)$ has atomless density, the LHS in the above inequality converges to 0 when τ goes to 0, and it is continuously increasing in τ. For any given $\epsilon > 0$, the RHS in the above inequality is strictly positive. Therefore, we can always find $\tilde{\tau} > 0$ such that $\mathcal{W}$ dominates $\mathcal{A}$ whenever $\tau < \tilde{\tau}$.

Thus, under a timely disaster alert, if $\theta \notin \Theta^L$, then, since all the agents wait, the disaster alert is not triggered, regardless of whether the shock has arrived or not. This means $\Theta^P(\Gamma^\tau) = \emptyset$ when $\tau < \tilde{\tau}$. ■

Timely disclosure is critical to this result. First, as in the benchmark model, waiting for a short period is not very costly conditional on the fact that the bank has not failed yet (since $u(t)$ is continuous). More importantly, when the shock arrives stochastically over time, a timely disclosure limits the chance that the shock arrives before the time of disclosure. Thus, it ensures that the disaster alert is an early warning with a sufficiently high probability. Based on this mechanism, theorem 3 demonstrates that when the shock arrives stochastically, a forward-looking weakest stress test eliminates panic as long as it is timely.

Bank Stress Test in Practice One of the most important functions of the stress test is to provide credible information about how banks are likely to perform under severely distressed macroeconomic conditions. In practice, the regulators collect data from the banks, and based on their models, investigate how the banks' balance sheets performs under different kinds of stress scenarios that may arise *in the future*. We demonstrate the importance of this forward-looking feature in averting financial panic. If it is known that the shock has already arrived, then the agents cannot be persuaded to wait for a test. However, if the agents do not know when the shock will arrive, then they consider a timely stress test very likely to be an early warning and wait for it rather than panicking.

The U.S. government responded to the aftermath of the financial panic during the great recession in 2008 with various measures such as liquidity injection and debt guarantees. Stress tests, which involve information production and disclosure, emerged as a potent tool to quell panic during times of economic uncertainty.[24] There is ample evidence that the U.S. stress tests produced credible information about the financial institutions and helped restore confidence in the banking system.[25] The contemporaneous stress tests in Europe were not as successful. While the tests in 2009 – 2010 did not have a noticeable impact, what truly stands out is the failure of Dexia, a bank that had to be bailed out three months after it had passed the 2011 stress test (See Acharya, Engle and Pierret (2014)). Anderson (2016) argues that this contrasting experiences can be attributed to the following: While

[24]In practice, a stress test policy is also supplemented by other measures of banking regulation, which is not a part of our model. For examples of such measures, see Faria-e Castro, Martinez and Philippon (2016) for fiscal capacity; Shapiro and Skeie (2015) for the cost of injecting capital; and Orlov, Zryumov and Skrzypacz (2018) for the co-determination of stress tests disclosure and capital requirements.

[25] Peristiani, Morgan and Savino (2010) documents evidence to show that stress tests helped quell the financial panic by producing vital information about banks. Bernanke (2013) admits *"Supervisors' public disclosure of the stress tests results helped restore confidence in the banking system..."* Gorton (2015) states that the tests results were viewed as credible, and the stress tests are widely viewed as a success.

the stress test attempt in the U.S. was "timely," it was perhaps, "too little, too late" in the Europe.[26] This paper provides the theoretical rationale behind this claim.

5 Discussion

We consider a canonical regime change game where agents' private signals could be arbitrarily correlated, and the principal does not have access to these private signals. She adopts a simple policy — a timely disaster alert — and she does not need "ex-ante" commitment to enforce such a policy. Recall that unlike the stress test proposed in Goldstein and Huang (2016) and Inostroza and Pavan (2018), in our proposed policy, the ex-post incentive compatibility is not violated. Yet, surprisingly, the policy completely eliminates panic. Even when new information arrives over time, a timely alert set for each time new information arrives stops the agents from panicking. We also saw how this insight can be used to design forward-looking stress tests.

In some application, the regulator may want more than just survival of the regime. For example, she may want minimum possible attack when the regime survives. Note that under the timely alert policy, there is no attack against a regime that survives. Thus, the principal achieves her objective. However, when the regime fails ($\theta \in \Theta^L$), all agents attack at time τ rather than at the beginning, and thus bear an unnecessary waiting cost. If the principal does not want the agents to bear such waiting costs, then she can supplement the timely disaster alert policy with a time zero disclosure policy that reveals whether $\theta \in \Theta^L$. It is also possible that the principal may want to delay the inevitable regime change. Then, without ex-ante commitment, the principal cannot implement a timely disaster alert policy.[27]

In reality, some of the features of our benchmark setup could be different. In this section, we discuss the role of these assumptions, and evaluate whether they are essential or could be relaxed.

[26] In the U.S., the plan for stress testing was announced on February 10, 2009. The white paper describing the procedures employed in SCAP was released on April 24, 2009 and the results of the SCAP were disclosed on May 7, 2009. On the other hand, although signs of instability in the financial system became apparent around the same time as in the U.S., the first European stress tests were conducted in October 2009.

[27] Orlov, Skrzypacz and Zryumov (2019) and Ely and Szydlowski (2019) consider a single long lived agent who decides when to quit and the principal wants him to quit later than earlier.

5.1 Sequential Rationality

We use sequential rationality to argue that an agent who has waited for the disaster alert will not attack after the alert is not triggered (option value lemma). If an agent believes that others are sequentially rational, then it follows from the option value lemma that he is not facing any strategic uncertainty after the alert. We leverage this fact to show that the agents will wait for the alert if it is set in a timely manner (timely alert lemma).

It is worth mentioning that in the epistemic game theory literature (see Dekel and Siniscalchi (2015) for a recent survey), the solution concept that uses the common knowledge of (sequential) rationality at the beginning of the game is called *initial rationalizability*. We use this solution concept in an incomplete information setting. This problem has an intuitive connection to the coordination with an outside option example in the forward induction literature (see Pearce (1984); Kohlberg and Mertens (1986); Van Damme (1989); Ben-Porath and Dekel (1992)). The forward induction argument requires agents to continue believing in other agents' sequential rationality even after seeing an unexpected history (see Battigalli and Siniscalchi (2002)). Given the fundamental uncertainty and the doubt assumption, it is always possible that the alert will not be triggered. Thus, neither the history $d^\tau = 0$ nor the history $d^\tau = 1$ is unexpected. Thus, any constraint on the beliefs conditional on unexpected histories, or the forward induction refinement, is unnecessary for our result.[28]

However, if a rational agent fears that others are not sequentially rational, then the strategic uncertainty after the alert may remain. Therefore, the agents may not want to wait for such an alert, and this may lead to panic. Suppose there is a chance η that, some agents who have waited will behave irrationally and attack even after the disaster alert is not triggered, and hence causing a regime change. The following proposition shows that if the probability η is sufficiently small, a timely disaster alert will eliminate panic.

Proposition 1 *Suppose even after the alert is not triggered the chance of regime change is η because of irrational attacks. If $\eta < \frac{g-1}{g-\underline{l}}$, then there exists a $\hat{\tau}^\eta > 0$ such that Γ^τ with $\tau < \hat{\tau}^\eta$ eliminates panic.*

Recall that while the cost of waiting for a timely alert can be arbitrarily small, there is a strictly positive benefit from waiting. If the agents are not certain about the sequential

[28] Also, unlike the coordination with an outside option example, in our setup, all the agents can attack early. When both agents have outside options (or attacking early in our setup), they face another coordination problem — whether to take the outside option or play the coordination game that follows.

rationality of the other agents, then this benefit is lower. However, if η is small enough, then the benefit still remains strictly positive. In this sense, the result is robust even if we relax the assumption of common knowledge of (sequential) rationality.

Suppose that the principal does not exactly know θ or N_τ. Consequently, even though the disaster alert has not been triggered, the regime may change with probability η. Thus, η arises from the incompetence of the principal rather than the irrationality of the agents. It is easy to see that by same argument as above, if the principal is competent enough in learning the state and the history, then a timely disaster alert will eliminate panic.

5.2 (Un)necessary Doubt

The policy of a timely disaster alert works with a relatively flexible set of exogenous information structure. The only restriction we impose on the exogenous information is the *doubt* assumption. Note that this assumption is stronger than saying that there is an upper dominance region. It requires that, regardless of the private signal, an agent believes that θ could be in the upper dominance region. This assumption can be weakened depending on the heterogeneity of agents' beliefs as the following example shows.

Example 2 *The regime change function $R(\theta, N) = \theta - N$. Nature draws θ from $\mathcal{U}[-1, 2]$, and agents receive independent private signals $s_i \in \{\mathsf{L}, \mathsf{M}, \mathsf{H}\}$ according to the following conditional distribution ($p \in (0.5, 1)$)*

$f_i(s_i \mid \theta)$	L	M	H
$\theta \in \Theta^L = [-1, 0)$	p	$(1-p)$	0
$\theta \in [0, 1)$	$\frac{1}{2}(1-p)$	p	$\frac{1}{2}(1-p)$
$\theta \in \Theta^U = [1, 2]$	0	$(1-p)$	p

In the above example 2, $f(s_i|\theta)$ does not have full support. If an agent receives the signal L, then the agent knows for sure that the fundamental of the regime is not in the upper dominance region, that is, $\theta < 1$. Hence, the *doubt* assumption is violated.

In this case, an agent receiving signal L knows that attacking immediately is not a mistake if others also attack immediately. Suppose all agents receive the public signal $s = \mathsf{L}$. Then, even under any timely disclosure policy Γ^τ, all agents attacking immediately is a possible equilibrium. Thus, under the general information structure, assumption 1 is indeed a necessary condition for our main result.

However, this does not mean that a given information structure must satisfy assumption 1 for the result to be true. Let us go back to example 2, but now suppose that the signals are not public. In particular, suppose that the signals are conditionally independent.

First, note that the agents who receive $s_i = \mathsf{M}$ or H believe there is a positive chance that the fundamental of the regime is in the upper dominance region, that is, $\mathbb{P}(\theta \geq 1|s_i = \mathsf{H}) = \frac{2p}{1+p} > \mathbb{P}(\theta \geq 1|s_i = \mathsf{M}) = \frac{1-p}{2-p}$. Hence, the *doubt* assumption holds for $s_i = \mathsf{H}, \mathsf{M}$ for any $\epsilon < \frac{1-p}{2-p}$, and agents receiving these two signals will not attack immediately (Lemma 2) under an appropriate timely disaster alert.

Now consider the agent who receives signal $s_i = \mathsf{L}$. He knows that the fundamental is not in the upper dominance region, and thus, if all other agents attack immediately, the alert will be triggered for sure. However, since signals are not public, he understands that some other agents may have received signal $s_{-i} = \mathsf{M}$ or H, and thus, avoid attacking. Only agents who have received $s_{-i} = \mathsf{L}$ may attack before the time of disclosure, and thus, the maximum attack is the share of agents with signal L. This means that the alert cannot be triggered if the realized fundamental θ is greater than the fraction of agents with signal L, that is, $\mathbb{P}(d^\tau = 0|\mathsf{L})$ is at least

$$\begin{aligned} & \mathbb{P}(\theta \geq \mathbb{P}(s_{-i} = \mathsf{L})|\theta \in [0,1))\mathbb{P}(\theta \in [0,1)|\mathsf{L}) \\ = \; & \mathbb{P}(\theta \geq \frac{1}{2}(1-p)|\theta \in [0,1)) \times \frac{1-p}{1+p} = \frac{1-p}{2}. \end{aligned}$$

This shows that even the agent who receives $s_i = \mathsf{L}$ believe that there is a strictly positive probability that the disaster alert will not be triggered. Thus, attacking immediately (strategy $\mathcal{A}$) might be a mistake; when the alert is set in a timely manner, the strategy of "wait and see" ($\mathcal{W}$) is the dominant one. [29]

Now suppose that the information structure is as follows: with probability α, $s_j = s_i$; and with probability $(1-\alpha)$, s_j is a conditionally independent signal. If $\alpha = 1$, then this captures the public information case, and if $\alpha = 0$, then this captures the conditionally independent signal case. A lower α indicates more heterogenous beliefs. For any given $\alpha < 1$,

$$\mathbb{P}(d^\tau = 0|\mathsf{L}) \geq (1-\alpha)\frac{1-p}{2}.$$

Thus, if α is away from 1, that is, there is enough heterogeneity in agents' beliefs, the

[29] Note that this argument requires that the agents with signal L should not only believe that others are rational, but also that the agents with signal M and H believe that others are rational and thus they will never attack before the disclosure.

principal can eliminate panic. In this sense, the doubt assumption is unnecessary.

5.3 (Ir)reversibility and Panic

The principal wants to dissuade agents from attacking, and attack is an irreversible action. The disaster alert at some pre-specified time τ discloses whether the irreversible action has become the dominant strategy. Suppose, in contrast to our benchmark set up, the principal wants to persuade agents to play the irreversible action. The analogue of the disaster alert in this case is to disclose whether the reversible action has become the dominant strategy. However, the option value argument only applies to an irreversible action. Take, for example, a regime change investment game (similar to Gale (1995)), where the investment is irreversible and the principal wants to persuade the agents to invest. Because not investing is reversible, unlike Lemma 1, we cannot argue that an agent who does not invest at time τ, will be better off if he does not invest at time 0. Thus, an essential feature of this model is that the principal wants to dissuade the agents from playing the irreversible action.

One may think that if the attack is also a reversible action, then the result would be stronger because the agents who stayed would be reassured by the agents who had left, but can come back. However, if the attack is also reversible, under the disaster alert policy, an agent can adopt the following strategy ($\mathcal{W}^c$): attack right away and reverse the action only if the alert is not triggered. This could trigger the alert, although it was not warranted, and thus panics is not eliminated. In the online appendix, we construct a simple example to show that when both attacking and not attacking are reversible, under the disaster alert policy, there exists an equilibrium in which agents attack the regime even though the failure of the regime is not warrented.

6 Conclusion

This study considers a dynamic regime change game with privately informed agents and proposes a dynamic information disclosure policy that completely eliminates panic. The policy is a timely alert that warns the agents if the regime is doomed to fail. This proposed policy is simple to implement. It does not require the information designer to have the access to agents' private signals. It is a binary and public disclosure. It does not require ex-ante commitment. Yet, the principal achieves the first-best outcome. We use the theoretical insight to build a forward-looking stress test that prevents bank runs. The literature shows that information manipulation cannot eliminate panic when agents move simultaneously. In contrast, this study shows that if there is a little time window and the principal can disclose information about the endogenous history, panic can be completely eliminated.

References

Acharya, Viral, Robert Engle, and Diane Pierret. 2014. "Testing macroprudential stress tests: The risk of regulatory risk weights." *Journal of Monetary Economics*, 65: 36–53.

Anderson, R. W. 2016. "Stress testing and macroprudential regulation: A transatlantic assessment." *CEPR Press*.

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): 711–756.

Au, Pak Hung. 2015. "Dynamic information disclosure." *The RAND Journal of Economics*, 46(4): 791–823.

Ball, Ian. 2019. "Dynamic Information Provision: Rewarding the Past and Guiding the Future." *Available at SSRN 3103127*.

Basak, Deepal, and Zhen Zhou. 2020. "Diffusing Coordination Risk." *American Economic Review*, 110.1: 271–297.

Battigalli, Pierpaolo, and Marciano Siniscalchi. 2002. "Strong belief and forward induction reasoning." *Journal of Economic Theory*, 106(2): 356–391.

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

Bergemann, Dirk, and Stephen Morris. 2016. "Bayes correlated equilibrium and the comparison of information structures in games." *Theoretical Economics*, 11.2: 487–522.

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

Carlsson, Hans, and Eric Van Damme. 1993. "Global Games and Equilibrium Eelection." *Econometrica*, 61(5): 989–1018.

Chamley, Christophe, and Douglas Gale. 1994. "Information revelation and strategic delay in a model of investment." *Econometrica: Journal of the Econometric Society*, 1065–1085.

Che, Yeon-Koo, and Johannes Hörner. 2018. "Recommender systems as mechanisms for social learning." *The Quarterly Journal of Economics*, 133(2): 871–925.

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

Dasgupta, Amil, Jakub Steiner, and Colin Stewart. 2012. "Dynamic Coordination with Individual Learning." *Games and Economic Behavior*, 74(1): 83–101.

Dekel, Eddie, and Marciano Siniscalchi. 2015. "Epistemic game theory." *Handbook of Game Theory with Economic Applications*, Vol 4: 619–702.

Doval, Laura, and Jeffrey Ely. 2019. "Sequential information design." *Working Paper*.

Edmond, Chris. 2013. "Information Manipulation, Coordination, and Regime Change." *Review of Economic Studies*, 80(4): 1422–1458.

Ely, Jeffrey, and Martin Szydlowski. 2019. "Moving the Goalposts." *Journal of Political Economy*, forthcoming.

Faria-e Castro, Miguel, Joseba Martinez, and Thomas Philippon. 2016. "Runs versus lemons: information disclosure and fiscal capacity." *The Review of Economic Studies*, 84(4): 1683–1707.

Gale, Douglas. 1995. "Dynamic Coordination Games." *Economic theory*, 5(1): 1–18.

Goldstein, Itay, and Chong Huang. 2016. "Bayesian Persuasion in Coordination Games." *The American Economic Review*, 106(5): 592–596.

Gorton, Gary. 2015. "Stress for success: A review of Timothy Geithner's financial crisis memoir." *Journal of Economic Literature*, 53(4): 975–95.

Gul, Faruk, and Russell Lundholm. 1995. "Endogenous timing and the clustering of agents' decisions." *Journal of political Economy*, , (103.5): 1039–1066.

Inostroza, Nicolas, and Alessandro Pavan. 2018. "Persuasion in Global Games with Application to Stress Testing." http://faculty.wcas.northwestern.edu/~apa522/persuasion-GG.pdf.

Kohlberg, Elon, and Jean-Francois Mertens. 1986. "On the Strategic Stability of Equilibria." *Econometrica*, 54(5): 1003–1037.

Kolb, Aaron, and Erik Madsen. 2019. "Leaks, Sabotage, and Information Design." *Available at SSRN 3327880*.

Li, Fei, Yangbo Song, and Mofei Zhao. 2019. "Global Manipulation by Local Obfuscation: Information Design in Coordination Games." *Available at SSRN*.

Mathevet, Laurent, and Jakub Steiner. 2013. "Tractable dynamic global games and applications." *Journal of Economic Theory*, 148(6): 2583–2619.

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)*. , ed. Dewatripont, Hansen and Turnovsky. Cambridge University Press.

Orlov, Dmitry, Andrzej Skrzypacz, and Pavel Zryumov. 2019. "Persuading the Principal To Wait." *Journal of Political Economy*, forthcoming.

Orlov, Dmitry, Pavel Zryumov, and Andrzej Skrzypacz. 2018. "Design of macro-prudential stress tests." *working paper*.

Pearce, David G. 1984. "Rationalizable strategic behavior and the problem of perfection." *Econometrica: Journal of the Econometric Society*, 1029–1050.

Peristiani, Stavros, Donald P Morgan, and Vanessa Savino. 2010. "The information value of the stress test and bank opacity." *FRB of New York Staff Report*, , (460).

Shapiro, Joel, and David Skeie. 2015. "Information management in banking crises." *The Review of Financial Studies*, 28(8): 2322–2363.

Van Damme, Eric. 1989. "Stable Equilbria and Forward Induction." *Journal of Economic Theory*, 48: 476–496.

Appendix

Proof of Theorem 2. For $k = 0, 1, 2, \ldots, K-1$, let $\mathcal{W}^k$ be the strategy in which an agent waits until the time $(t_k + \tau)$ for the $(k+1)$th disaster alert and follows all the alerts along the path, that is, attacks only when $d^{t_{k'}+\tau} = 1$ for all $k' \leq k$, but does not wait for the next disaster alert, and attacks at time t_{k+1}. Under the strategy $\mathcal{W}^K$, agents wait for all disaster alerts and follow them.

Lemma 3 *Under the extended disclosure policy Γ^τ, for any noisy signal $\{s_i^k\}_{k=0}^K$, the only rationalizable strategies are $\mathcal{A}$ and $\{\mathcal{W}^k\}_{k=0}^K$.*

Proof. Because of the delay cost, attacking at any time when there is no new arrival of information is strictly dominated by attacking at an earlier time with the same information. Hence, attack can only happen at time t_k or $t_k + \tau + dt$ for $k = 0, 1, \ldots, K$. The agents who have waited for the $(k+1)$th disaster alert will attack right away if $d^{t_k+\tau} = 1$. It follows from the option value argument that if $d^{t_k+\tau} = 0$, an agent would not attack regardless of $\{s_i^{k'}\}_{k' \leq k}$. Otherwise, the new information disclosed at $t_k + \tau$ does not have any value and the agent should attack earlier without waiting for the new information. Hence, each rationalizable strategy can be described as either wait until the $(k+1)$th disaster alert for $k = 0, 1, 2, \ldots, K$, or do not wait for any disaster alert at all and attack right away, that is, $\mathcal{A}$. ■

Lemma 4 *Given generalized doubt assumption, under the extended disclosure policy Γ^τ, where $\tau < \min\{\hat{\tau}, \bar{\tau}\}$, for any signal realization $\{s_i^k\}_{k=0}^K$, the only rationalizable strategy for an agent is $\mathcal{W}^K$, i.e., wait for all disaster alert and follow the principal's recommendation.*

Proof. Under any rationalizable strategy, if any disaster alert is triggered before time t_K, all agents attack before t_K. Now, suppose the disaster alerts have not been triggered before time t_K, that is, $d^{t_{K-1}+\tau} = \ldots = d^{t_0+\tau} = 0$. For the agents who have waited until the Kth disaster alert (disclosed at time $t_{k-1} + \tau < t_k$), attacking at t_k without waiting for the $(K+1)$th alert (strategy $\mathcal{W}^{K-1}$) is strictly dominated by waiting for the next alert and following it (strategy $\mathcal{W}^K$) when $\tau < \hat{\tau}$, regardless of the private signal s_i^K and what other agents do (same argument as in Lemma 2). Hence, regardless of which rationalizable strategy agents take, if the previous alert was not triggered, there is no attack between the

Kth alert and the $(K+1)$th alert, and hence, $d^{t_K+\tau}=0$. Otherwise, if $d^{t_{K-1}+\tau}=1$, then $d^{t_K+\tau}=1$.

Therefore, under timely alerts ($\tau < \hat{\tau}$) , no rational agents would act on their new information arriving at t_K, and thus $d^{t_K+\tau} = d^{t_{K-1}+\tau}$. By repeating the same argument again, it is easy to prove that $\mathcal{W}^K$ strictly dominates $\mathcal{W}^k(k=0,1,\ldots,K-1)$ and $\mathcal{A}$. ■

Hence, under the extended disclosure policy Γ^τ with $\tau < \min\{\hat{\tau}, \bar{\tau}\}$ and the unique rationalizable strategy $\mathcal{W}^K$, agents never attack a regime with $\theta \notin \Theta^L$ regardless of their signals. Following the same argument as in the proof of Theorem 1, $\Theta^P(\Gamma^\tau)=\emptyset$. □

Proof of Proposition 1. Consider an agent with signal s_i. The net payoff from playing $\mathcal{W}$ compared with $\mathcal{A}$, $D(\Gamma^\tau, s_i)$, is

$$\mathbb{P}(d^\tau=1|s_i)(u(\tau)-u(0))+\mathbb{P}(d^\tau=0|s_i)\left(\mathbb{E}\left[(1-\eta)g(\theta,N_T)+\eta l(\theta,N_T)|s_i\right]-1\right).$$

Using Lipschitz continuity of $u(t)$ and assumption 1, we have

$$D(\Gamma^\tau,s_i)\geq -(1-\epsilon)\mathcal{C}\tau+\epsilon((1-\eta)(\underline{g}-1)+\eta(\underline{l}-1)).$$

If $\eta < \frac{\underline{g}-1}{\underline{g}-\underline{l}}$, then a timely disaster alert policy Γ^τ, where $\tau < \hat{\tau}^\eta \equiv \frac{\epsilon}{1-\epsilon}\frac{(1-\eta)(\underline{g}-1)-\eta(\underline{1}-l)}{\mathcal{C}}$, eliminates panic. □

Online Appendix

Reversibility

In the following example, we consider a regime change game in which both actions are reversible, and the payoff depends on the duration of the actions.

Example 3 *The investors decide whether to stay in (not attack) or stay out (attack). They can make the switch at any point with in the time window $[0, T]$ since both actions are reversible. The regime change can only happen in the end of the time window, governed by the function $R(\theta, N) = \theta - N$.*

The payoffs depend on the duration of each action. If an investor stays in for a duration of $t \in [0, T]$, then he receives a higher flow return (g) or lower one (l), depending on the fate of the regime, weighted by the duration that he stays in. For the duration that he stays out ($T - t$), he receives a fixed flow return r. Thus, he obtains

$$(g \cdot \mathbb{1}\{\theta \geq N_T\} + l \cdot \mathbb{1}\{\theta < N_T\})\, t + r(T - t),$$

where $g > r > l$.

Regarding the fundamental strength of the regime θ, investors have a common improper prior ($\theta \sim \mathcal{U}[\mathcal{R}]$) and private information with conditionally independent noise, that is, $s_i = \theta + \sigma\varepsilon_i$, where $\varepsilon_i \sim N(0, 1)$ and $\sigma > 0$ scales the random noise ε_i. Investors cannot learn about others' actions within the time window.

Suppose the principal sets a timely disaster alert $\mathit{\Gamma}^\tau$: $d^\tau = 1$ if and only if $\theta < N_\tau$. When attack is also reversible, an agent can always take the following strategy: exit at time 0 and return at $\tau + dt$ if the alert is not triggered. If all the agents take this strategy, then the alert will be triggered unless $\theta \geq \overline{\theta}$. If an agent believes that θ is unlikely to be in the upper dominance region, and others will follow this above strategy, then he will do the same. We call this strategy $\mathcal{W}^c$. Clearly, this is a rationalizable strategy. Below, we construct an equilibrium in which the agents panic regardless of however timely disaster alert.

Proposition 2 *Under the disaster alert policy $\mathit{\Gamma}^\tau$, there exists an equilibrium in which investors with $s_i < s^*$ play $\mathcal{W}^c$ and the ones with $s_i \geq s^*$ play $\mathcal{W}$, regardless of the disclosure time τ. Accordingly, for $\theta \in (0, \frac{1}{1+\frac{g-r}{r-l}})$, the regime changes, although it is not warranted.*

Proof. Let us define a cutoff strategy $\hat{s}$ as follows: the agent plays $\mathcal{W}^c$ if $s_i < \hat{s}$ and plays $\mathcal{W}$ if $s_i \geq \hat{s}$. To construct an equiibrium with such a cutoff strategy, we want to find a threshold s^* such that the cutoff strategy s^* is the best response of any agent i when all others are playing the same cutoff strategy.

Suppose all other agents play a cutoff strategy $\hat{s}$. This means that all of them will stay in after seeing $d^\tau = 0$, and thus the regime does not change in the end. On the other hand, all of them will stay out after seeing $d^\tau = 1$ and thus the regime changes in the end. That, in turn, implies that agent i would stay out after seeing $d^\tau = 1$ and stay in otherwise. The only remaining problem for agent i is whether he should stay in or not before the disclosure, and thus strategy $\mathcal{W}^c$ (staying out until τ) and $\mathcal{W}$ (staying in until τ) are the only possible strategies he would take.

Given $\hat{s}$, for any θ, the aggregate attack (from agents who play $\mathcal{W}^c$) at time 0 is

$$\mathbb{P}(s_i < \hat{s}|\theta) = \Phi\left(\frac{\hat{s}-\theta}{\sigma}\right).$$

Clearly, it is decreasing in θ. Therefore, there exists a $\hat{\theta}$ such that the alert is triggered ($d^\tau = 1$) if and only if $\theta < \hat{\theta}$, where

$$\Phi\left(\frac{\hat{s}-\hat{\theta}}{\sigma}\right) = \hat{\theta}.$$

An agent with private signal s_i, believes that the net expected payoff from playing $\mathcal{W}^c$ compared to $\mathcal{W}$ is

$$\left(1-\Phi\left(\frac{s_i-\hat{\theta}}{\sigma}\right)\right)(r-l)\tau + \Phi\left(\frac{s_i-\hat{\theta}}{\sigma}\right)(r-g)\tau.$$

Clearly, it is decreasing in s_i. Hence, for the cutoff strategy $\hat{s}$ to be the best response, we only need to verify that agent i is indifferent between $\mathcal{W}^c$ and $\mathcal{W}$ when he gets the threshold signal $\hat{s}$. Substituting s_i with threshold signal $\hat{s}$ and plugging $\Phi((\hat{s}-\hat{\theta})/\sigma) = \hat{\theta}$ into the payoff difference, we get

$$\Big((r-l) - \hat{\theta}((g-r)+(r-l)\Big)\tau = 0 \implies \hat{\theta} = \frac{1}{1+\frac{g-r}{r-l}}.$$

Thus, regardless of τ, $s^* = \hat{\theta} + \frac{1}{\sigma}\Phi^{-1}(\hat{\theta})$ constitutes an equilibrium, and in this equilibrium, a regime with the fundamental below $\hat{\theta}$, but above 0, will not survive because of panics. ■

Tax on capital flight

Consider the capital outflow example. However, unlike our benchmark setup, suppose the investors receive flow payoff from staying. If an investor has not exited by time $t \in [0, T]$, he receives a flow payoff at time t depending on the underlying fundamental θ and aggregate exit so far N_t as follows.

$$\tilde{\rho}(\theta, N_t) = \begin{cases} \overline{\rho} & \text{if } R(\theta, N_t) \geq 0 \\ \underline{\rho} & \text{if } R(\theta, N_t) < 0. \end{cases}$$

If $R(\theta, N_T) \geq 0$ (the regime survives), then the investor who does not exit will earn a flow payoff $\overline{\rho}$ forever. However, if $R(\theta, N_t)$ becomes less than 0 at some t, then the investor who does not exit will start receiving flow payoff $\underline{\rho}$ from time t onward. Thus, if the investor stays, then his payoff is

$$v(\theta, (N_t)) = \int_{t=0}^{T} e^{-\beta t} \tilde{\rho}(\theta, N_t) dt + \int_{T}^{\infty} e^{-\beta t} \tilde{\rho}(\theta, N_T) dt,$$

where $\beta > 0$ is the discount rate. On the other hand, if an investor exits at some time t, and then switches to a safe investment project, it yields a fixed flow return of $\rho > 0$, where

$$\overline{\rho} > \rho > \underline{\rho}.$$

Note that, unlike the benchmark setup, the regime can change at any time t, than only at T. The flow payoff acts as endogenous disaster alert: Whenever the flow payoff from keeping the investment at the emerging market becomes $\underline{\rho}$, the agents learn that the regime has changed. Unlike our benchmark setup, a delayed attack may not be costly, since the investor could earn $\overline{\rho}$ rather than ρ by exiting later.

The government in emerging economies often imposes a tax on capital flight. Clearly, a tax would discourage investors from exiting the market. However, this could discourage them to invest in the first place (see Mathevet and Steiner (2013)). We make a different argument. A tax on capital flight makes waiting costly, that is, if an investor exits, then he should exit earlier than later. Suppose that the investor has to pay a tax at a rate $\mu \in (0, 1)$

on the flow payoff the investor has earned until time t. Thus, his payoff from withdrawing at time $t_i \in [0, T]$ is

$$u(\theta, t_i, (N_t)) \equiv \int_{t=0}^{t_i} e^{-\beta t}(1-\mu)\tilde{\rho}(\theta, N_t)dt + \int_{t_i}^{\infty} e^{-\beta t}\rho dt.$$

Assumption 2 *(1)* $\mu > 1 - \frac{\rho}{\bar{\rho}}$ *and (2)* $T < \frac{1}{\beta}\ln(1 + \frac{\rho - \underline{\rho}}{\bar{\rho}})$.

The first restriction in Assumption 2 ensures that the tax rate is high enough such that the the flow payoff from safe outside option dominates the after-tax flow payoff from staying, that is, $(1-\mu)\overline{\rho} < \rho$. This implies that u is decreasing in t, and thus, if an investor exits, he should exit as early as possible. The second restriction in Assumption 2 ensures that the time window is sufficiently small. Otherwise, an investor may accumulate significant flow payoff over time and may not want to exit, because of the significant exit tax.

Proposition 3 *In the capital outflow game, under Assumption 2, the investors do not panic. That is,* $\Theta^P = \emptyset$.

Proof. Suppose that the flow payoff switches from $\bar{\rho}$ to $\underline{\rho}$ at time τ. Then, the expected payoff from waiting and not exiting at $\tau + dt$ is

$$v(\theta, (N_t)_{t<\tau}) = \int_{t=0}^{\tau} e^{-\beta t}\tilde{\rho}(\theta, N_t)dt + \int_{t=\tau}^{\infty} e^{-\beta t}\underline{\rho}dt,$$

while the expected payoff from exiting right away is

$$u(\tau, \theta, N_\tau) = \int_{t=0}^{\tau} e^{-\beta t}(1-\mu)\tilde{\rho}(\theta, N_t)dt + \int_{t=\tau}^{\infty} e^{-\beta t}\rho dt.$$

So, the net payoff from waiting as compared to exiting right away is

$$\int_{0}^{\tau} e^{-\beta t}\mu\tilde{\rho}(\theta, N_t)dt - e^{-\beta\tau}\frac{\rho - \underline{\rho}}{\beta} \leq \frac{1}{\beta}\left[\mu\overline{\rho} - e^{-\beta\tau}(\mu\overline{\rho} + \rho - \underline{\rho})\right].$$

Note that the above inequality holds true for any possible $\tilde{\rho}$ before time τ (since $\tilde{\rho} \leq \bar{\rho}$). Based on the above inequality, one can easily check that, for any possible tax rate $\mu \in (0, 1)$, as long as $\tau < \frac{1}{\beta}\ln\frac{\rho - \underline{\rho} + \bar{\rho}}{\bar{\rho}}$, $v(\theta, (N_t)_{t<\tau}) < u(\tau, \theta, N_\tau)$ and thus investors would strictly prefer to exit immediately. Hence, if the second part of Assumption 2 holds true, $T < \frac{1}{\beta}\ln\frac{\rho - \underline{\rho} + \bar{\rho}}{\bar{\rho}}$, independent of the history (N_t) and the tax rate μ, investors would immediately exit once observing $\tilde{\rho} = \underline{\rho}$ at any point within the time window $[0, T]$.

It follows from the first part of assumption 2 that $u(\theta, t, N)$ is decreasing in t. Therefore, the option value argument holds true (lemma 1). Let us define $\mathcal{T} = \{t \in [0, T] | \rho(\theta, N_t) = \underline{\rho}\}$. If $\mathcal{T} = \emptyset$, then the switch does not happen, otherwise, let us write the first time when the flow payoff switches to $\underline{\rho}$ as $t_\rho := \min \mathcal{T}$. With the option value argument, the only ratonalizable strategies are (1) $\mathcal{A}$: exiting at time 0; and (2) $\mathcal{W}$: waiting and only exiting at $t_\rho + dt$ (not exiting if $\mathcal{T} = \emptyset$). Under these two strategies, the switch can only happen at time $0 + dt$. With the *doubt* assumption, the expected payoff from $\mathcal{A}$ is $\frac{\rho}{\beta}$, while that from $\mathcal{W}$ is

$$\epsilon(\frac{\bar{\rho}}{\beta}) + (1 - \epsilon)(\frac{1 - e^{-\beta dt}}{\beta}\bar{\rho} + \frac{e^{-\beta dt}}{\beta}\rho).$$

Clearly, $\mathcal{W}$ strictly dominates $\mathcal{A}$. Hence, if $\theta \notin \Theta^L$, the investors will not panic and start exiting. ■

It follows from the second part in Assumption 2 that after seeing $\tilde{\rho}(\theta, N_t) = \underline{\rho}$, an investor will exit right away. Because delay is costly, which follows from the assumption on the tax rate μ, any agent who did not exit early would only exit after seeing $\tilde{\rho} = \underline{\rho}$ (lemma 1). Since, the disaster alert is continuously in pace, it follows from lemma 2 that all the agents who believe that the flow payoff in future can be $\bar{\rho}$ with positive probability regardless of the other agents' actions (Assumption 1) would never exit unless $\tilde{\rho} = \underline{\rho}$ is realized. Hence, under such a capital taxation policy, there is no panic, or strategic capital outflows.