(In)Effectiveness of Signal Jamming in Regime-Change Games

Panagiotis Kyriazis   Edmund Y. Lou

July 10, 2025

Motivation

- Info manipulation/Signal-Jamming is a common practice
- Manipulation/Accounting Reporting Choices by/of banks:
  - Wells Fargo's fake accounts scandal
  - Deutsche Bank's hiding of losses
  - "Window dressing" or "selective reporting"
- Autocratic Regimes' use of propaganda
- Is it effective if observer rationally accounts for it (in equilibrium)?

Main Question

- Global Games deliver two thresholds on the fundamental (θ^*_{NM}) and agents' signals (x^*_{NM}) such that
  - Agents attack if and only if $x < x^*_{NM}$
  - Regime fails if and only if $\theta < \theta^*_{NM}$

Main Question

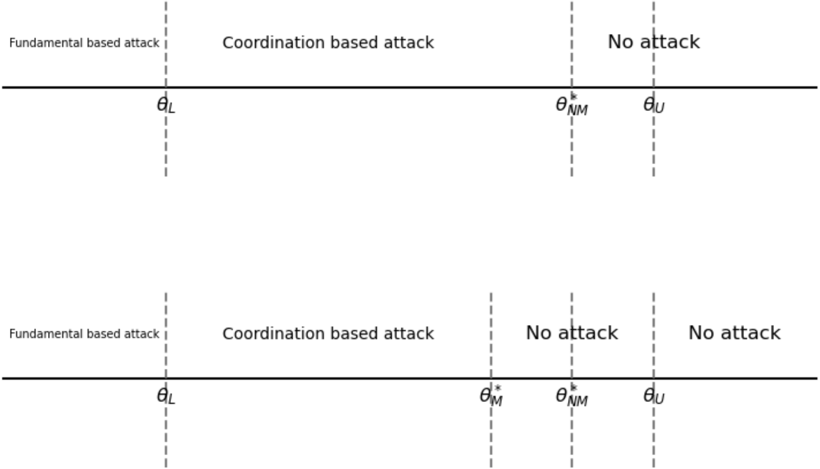
- Global Games deliver two thresholds on the fundamental (θ^*_{NM}) and agents' signals (x^*_{NM}) such that
  - Agents attack if and only if $x < x^*_{NM}$
  - Regime fails if and only if $\theta < \theta^*_{NM}$
- Signal-Jamming: regime can manipulate information at cost, agents do not observe regime's manipulation

Main Question

- Global Games deliver two thresholds on the fundamental (θ^*_{NM}) and agents' signals (x^*_{NM}) such that
  - Agents attack if and only if $x < x^*_{NM}$
  - Regime fails if and only if $\theta < \theta^*_{NM}$
- Signal-Jamming: regime can manipulate information at cost, agents do not observe regime's manipulation

Does the ability of the regime to jam signals improve the chances of survival, or, in other words, does it reduce the threshold below which regimes fail??

Main Question



Existing Results on Signal Jamming

- Distortion of actions by senders is well studied
- In many cases distortions have no systematic effect on the actions taken by the receiver

Existing Results on Signal Jamming

- Distortion of actions by senders is well studied
- In many cases distortions have no systematic effect on the actions taken by the receiver
- Symmetric info as in Holmström (1999) and Fudenberg and Tirole (1986)
    - Sender does not have private info about state so receiver perfectly infers signal distortion in equilibrium and can reverse it

Existing Results on Signal Jamming

- Distortion of actions by senders is well studied
- In many cases distortions have no systematic effect on the actions taken by the receiver
- Symmetric info as in Holmström (1999) and Fudenberg and Tirole (1986)
    - Sender does not have private info about state so receiver perfectly infers signal distortion in equilibrium and can reverse it
- Sender is privately informed but there is separating equilibrium
    - Receiver perfectly learns in equilibrium the info corresponding to the signal received, and can reverse the distortion (Rogoff and Sibert (1988))

Preview of Results

- Receiver will not be able to retrieve sender's private information from signal if:
    - It is received with noise
    - Pooling equilibrium

Preview of Results

- Receiver will not be able to retrieve sender's private information from signal if:
  - It is received with noise
  - Pooling equilibrium
- Signal jamming may be ineffective despite regime being perfectly informed

Preview of Results

- Receiver will not be able to retrieve sender's private information from signal if:
    - It is received with noise
    - Pooling equilibrium
- Signal jamming may be ineffective despite regime being perfectly informed
    - Ineffective when cost of attacking low
    - Effective when cost of attacking intermediate
    - Fully effective in the limit when signal precision is arbitrarily high

Plan

- Model
- Preliminary Analysis
- Equilibrium w/ Linear Manipulation Cost
- Conclusions

Model

Model

- Players
  - Continuum of agents, $i \in [0, 1]$
  - Regime
- State
  - Regime's type (fundamentals)
  - $\theta \sim$ improper prior
- Information
  - Agent i: $x_i = r + \varepsilon_i / \sqrt{\alpha}$
    - $r = \theta + m$: "report" sent by regime
    - $\varepsilon_i \sim_{IID} F$ w/ density f, which is symmetric about 0, log-concave, differentiable, strictly increasing on $(-\infty, 0)$

Model

- Choices
  - Agents:
    - Two actions:
      – Attack the status quo: $a_i = 1$
      – Refrain from attacking: $a_i = 0$
    - Aggregate action: $A = \int_{[0,1]} a_i \mathrm{d}i$
  - Regime:
    - Choose a report $r = \theta + m$
    - $m \geq 0$: signal jamming

Model

- Regime's payoff:

$$V(m;\theta,A) = \begin{cases} R(\theta,A) - C(m) & \text{if } R(\theta,A) \geq 0 \\ -C(m) & \text{if } R(\theta,A) < 0 \end{cases}$$

  - $R(\theta,A) = g(\theta) - A$: regime outcome function
    - $g(\theta)$: continuous, strictly increasing; $\exists \underline{\theta} < \overline{\theta}$: $g(\underline{\theta}) = 0$, $g(\overline{\theta}) = 1$
    - Status quo abandoned (regime change occurs) iff $R(\theta,A) < 0$
  - $C(m)$: cost of manipulating info
    - $C'(m) > 0$, $C''(m) \geq 0$
    - $C'(0) = k > 0$

Model

- Agents' payoff:

$$u(\theta, A) = \begin{cases} Z(\theta) - t & \text{if } R(\theta, A) < 0 \\ -t & \text{if } R(\theta, A) \geq 0 \end{cases}$$

  - $t > 0$: cost of attacking
  - $Z(\theta) > t$: continuous
- Solution concept:
  - Perfect Bayesian equilibrium (PBE)

Examples

- Rochet and Vives (2004) bank run model
- Morris and Shin (1999) currency attack model
- Edmond (2013) revolution model

Plan

- Model
- **Preliminary Analysis**
- Equilibrium w/ Linear Manipulation Cost
- Conclusions

Preliminary Analysis

Equilibrium w/o signal jamming

Proposition

Unique equilibrium is characterized by two thresholds θ^*_{NM} and x^*_{NM}:

(i) Regime change occurs iff $\theta < \theta^*_{\mathrm{NM}}$;

(ii) An agent attacks iff $x_i < x^*_{\mathrm{NM}}$.

Our definition

Definition

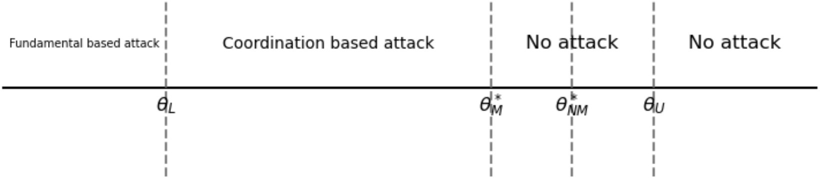
Suppose the regime jams agents' signals in equilibrium, then:

(i) Signal jamming is *effective* if $\theta^*_{\mathrm{M}} < \theta^*_{\mathrm{NM}}$;

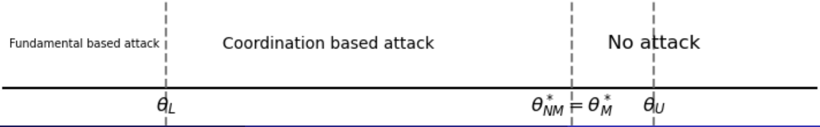
(ii) Signal jamming is *ineffective* if $\theta^*_{\mathrm{M}} = \theta^*_{\mathrm{NM}}$.

Our definition

Effective: Enabling more regime types to survive



Ineffective: Same regimes survive



Plan

- Model
- Preliminary Analysis
- **Equilibrium w/ Linear Manipulation Cost**
- Conclusions

Equilibrium w/ Linear Manipulation Cost

Regime's problem

- $C(m) = k \cdot m$
- Suppose all agents follow x^*_{M}
- Aggregate attack: $A(m;\theta) = F[\sqrt{\alpha}(x^*_{\mathrm{M}} - \theta - m)]$
- Regime θ's signal jamming policy:

$$m(\theta) \in \arg\max_{m \geq 0} V(m;\theta, A(m;\theta))$$

where

$$V(m;\theta, A(m;\theta)) = \begin{cases} \underbrace{R(\theta, A(m;\theta)) - C(m)}_{=\tilde{V}(m;\theta):\ \textit{manipulation payoff}} & \text{if } R(\theta, A(m;\theta)) \geq 0 \\ -C(m) & \text{if } R(\theta, A(m;\theta)) < 0 \end{cases}$$

Regime's problem

- No incentive to jam signals iff $\tilde{V}(\tilde{m}(\theta);\theta) < 0$, where

$$\tilde{m}(\theta) \in \arg\max_{m \geq 0} \tilde{V}(m;\theta)$$

- Equivalently,

$$\tilde{r}(\theta) = \tilde{m}(\theta) + \theta \in \arg\max_{r \geq \theta} \tilde{V}(r-\theta;\theta)$$

- FOC of relaxed problem (assuming $\alpha > (k/f(0))^2$):

$$\underbrace{\sqrt{\alpha} f[\sqrt{\alpha}(r - x^*_{\mathrm{M}})]}_{\text{marginal benefit}} = k \quad \Longrightarrow \quad r = x^*_{\mathrm{M}} + \underline{\rho}$$

Regime's problem

Lemma (partial pooling)

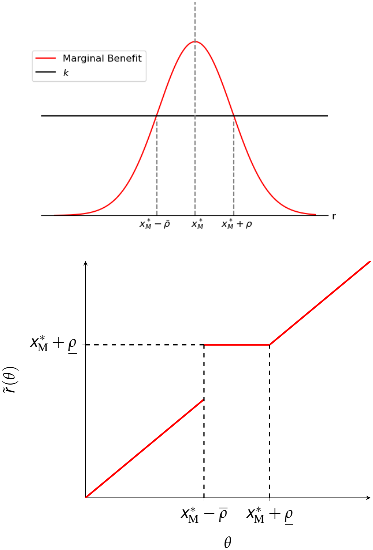
There exists $\overline{\rho} > 0$ such that

$$\tilde{r}(\theta) = \begin{cases} \theta & \text{if } \theta \notin [x^*_{\mathrm{M}} - \overline{\rho}, x^*_{\mathrm{M}} + \underline{\rho}] \\ x^*_{\mathrm{M}} + \underline{\rho} & \text{if } \theta \in [x^*_{\mathrm{M}} - \overline{\rho}, x^*_{\mathrm{M}} + \underline{\rho}] \end{cases},$$

where $\overline{\rho} > \underline{\rho}$ and solving

$$F(\sqrt{\alpha}\overline{\rho}) - k\overline{\rho} = F(-\sqrt{\alpha}\underline{\rho}) + k\underline{\rho}$$

Regime's problem



Equilibrium

- Agents attack if and only if expected payoff greater than cost
- Regime abandons status quo if and only if $g(\theta) < A$

Equilibrium pinned down by

$$g(\theta_M^*) = F(\sqrt{\alpha}(x_M^* - \theta_M^*))$$

$$\mathbb{E}_{\theta \sim \Psi(\cdot; x_i, m(\cdot))}[\mathbb{1}(\theta < \theta_{\mathrm{M}}^*) Z(\theta)] = t$$

where

$$\Psi(\theta; x_i, m(\cdot)) = \frac{\int_{-\infty}^{\theta} \sqrt{\alpha} f[\sqrt{\alpha}(x_i - \eta - m(\eta))] \mathrm{d}\eta}{\int_{-\infty}^{\infty} \sqrt{\alpha} f[\sqrt{\alpha}(x_i - \eta - m(\eta))] \mathrm{d}\eta}$$

is the agents' posterior CDF about θ

Equilibrium

Proposition

Unique equilibrium is characterized by two thresholds θ^*_{M} and x^*_{M}:

(i) Regime change occurs iff $\theta < \theta^*_{\mathrm{M}}$;

(ii) An agent attacks iff $x_i < x^*_{\mathrm{M}}$.

Effectiveness of signal jamming

Proposition

There exist two critical bounds of the cost of attacking the regime $\underline{t}$ and $\bar{t}$, such that $\underline{t} < \bar{t}$ and

(i) signal jamming is ineffective with $\theta^*_{\mathrm{M}} = \theta^*_{\mathrm{NM}}$ if $t \leq \underline{t}$;

(ii) signal jamming is effective with $\theta^*_{\mathrm{M}} < \theta^*_{\mathrm{NM}}$ if $\underline{t} < t < \bar{t}$;

(iii) signal jamming does not occur (i.e, $\theta^*_{\mathrm{M}} = \theta^*_{\mathrm{NM}}$) if $t \geq \bar{t}$.

Effectiveness of signal jamming

Proposition

There exist two critical bounds of the cost of attacking the regime $\underline{t}$ and $\overline{t}$, such that $\underline{t} < \overline{t}$ and

(i) signal jamming is ineffective with $\theta^*_{\mathrm{M}} = \theta^*_{\mathrm{NM}}$ if $t \leq \underline{t}$;

(ii) signal jamming is effective with $\theta^*_{\mathrm{M}} < \theta^*_{\mathrm{NM}}$ if $\underline{t} < t < \overline{t}$;

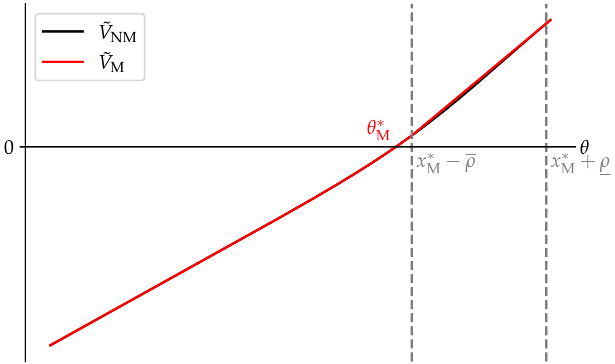
(iii) signal jamming does not occur (i.e, $\theta^*_{\mathrm{M}} = \theta^*_{\mathrm{NM}}$) if $t \geq \overline{t}$.

Moreover, manipulation happens on the interval

- $(x^*_M - \overline{\rho}, x^*_M + \underline{\rho})$ if $t \leq \underline{t}$
- $(\theta^*_{\mathrm{M}}, x^*_M + \underline{\rho})$ if $\underline{t} < t < \overline{t}$
- $\emptyset$ if $t \geq \overline{t}$

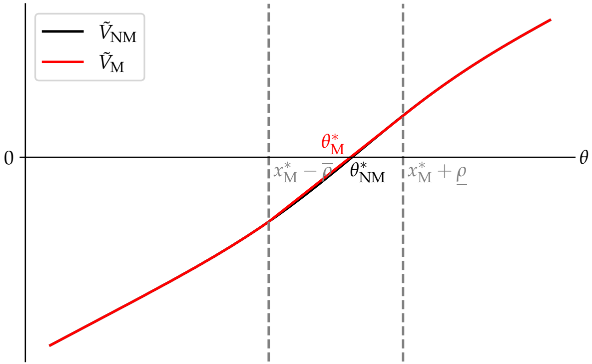
Ineffective signal jamming

If $t \leq \underline{t}$



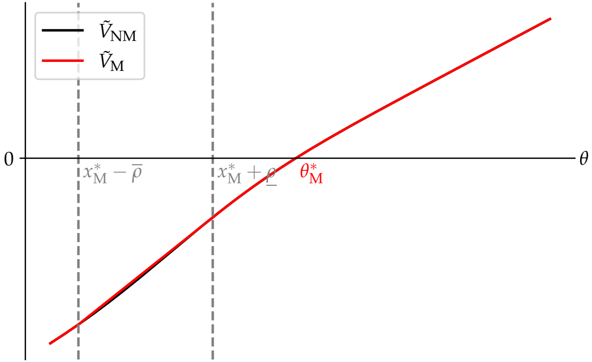
Effective signal jamming

If $\underline{t} < t < \bar{t}$



No signal jamming

If $t \geq \bar{t}$



Intuition: A Critical Result

Definition

The *belief distortion ratio* of the marginal agent is

$$\text{BDR} := \frac{\Psi\left(\theta^*_{\mathrm{M}} \mid x^*_{\mathrm{M}}, \mathrm{id}(\cdot)\right)}{\Psi\left(\theta^*_{\mathrm{M}} \mid x^*_{\mathrm{M}}, m(\cdot)\right)} = \int_{-\infty}^{\infty} \sqrt{\alpha} f\left[\sqrt{\alpha}(x^*_M - \theta - m(\theta))\right] d\theta.$$

- The BDR measures the relative change in marginal agent's belief about prob of regime change as a result of regime's info manipulation
- If BDR < 1, marginal agent perceives regime change as more likely than it actually is
- When BDR $= 1$, info manipulation does not affect belief

Intuition: A critical Result

Lemma

If $t \notin (\underline{t}, \bar{t})$, then BDR $= 1$; otherwise, BDR < 1.

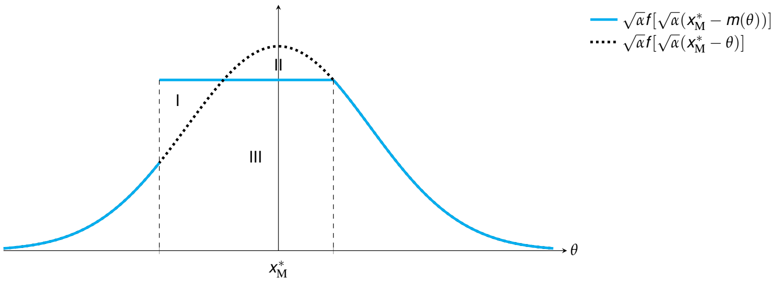


Figure: The regime's signal-jamming policy $m(\theta)$ ensures that the area of region I is equal to the area of region II when $t < \underline{t}$.

Effectiveness in the Limit

Proposition

The following hold: $\lim_{\alpha \to \infty} \underline{t} = 0$ and $\lim_{\alpha \to \infty} \bar{t} = \infty$. Moreover,

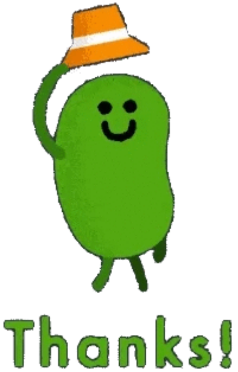
$$\lim_{\alpha \to \infty} \theta_M^* = 0$$

- As the signal's precision grows arbitrarily large (so only strategic uncertainty remains) signal jamming becomes fully effective.

Conclusions

Conclusions and Future Work

- (In)Effectiveness of signal jamming in regime-change games
  - Effective when agents' cost is intermediate
  - Ineffective (but occurs) when agents' cost is small
  - Fully effective in the limit when precision becomes arbitrarily high
- Future: Convex manipulation cost with
  - $c'(0) = 0$ (conjecture: always effective)
  - $c'(0) > 0$ (conjecture: ineffective)


Thanks!