Mechanisms for a No-Regret Agent: Beyond the Common Prior

Modibo Camara

Jason Hartline

Aleck Johnsen

Northwestern University

Cornell ORIE Seminar

Appeared in the 2020 Symposium on Foundations of Computer Science (FOCS)

Introduction

Motivation

Policy success or failure often influenced by...

- Environmental details (e.g. consumer demand, labor supply)
- Individuals' beliefs about environment (e.g. inflation expectations)

Inherently dynamic

- Economic conditions evolve in unpredictable ways
- Individuals learn and adapt over time

Ideal policies would adapt to environment over time

- E.g. minimum wage adapted to current, local market conditions

Motivation

Policy success or failure often influenced by...

- Environmental details (e.g. consumer demand, labor supply)
- Individuals' beliefs about environment (e.g. inflation expectations)

Inherently dynamic

- Economic conditions evolve in unpredictable ways
- Individuals learn and adapt over time

Ideal policies would adapt to environment over time

- E.g. minimum wage adapted to current, local market conditions

Motivation

Policy success or failure often influenced by...

- Environmental details (e.g. consumer demand, labor supply)
- Individuals' beliefs about environment (e.g. inflation expectations)

Inherently dynamic

- Economic conditions evolve in unpredictable ways
- Individuals learn and adapt over time

Ideal policies would adapt to environment over time

- E.g. minimum wage adapted to current, local market conditions

High Level Questions

Can we develop dynamic policies that adapt to their environment over time?

Without making assumptions on how the environment evolves?

- I.e. prior-free or adversarial

With permissive assumptions on agent behavior?

- Ex-ante optimal: Bayesian agents *want to* satisfy assumptions
- Ex-post feasible: non-Bayesian algorithm guaranteed to satisfy assumptions

High Level Questions

Can we develop dynamic policies that adapt to their environment over time?

Without making assumptions on how the environment evolves?

- I.e. prior-free or adversarial

With permissive assumptions on agent behavior?

- Ex-ante optimal: Bayesian agents *want to* satisfy assumptions
- Ex-post feasible: non-Bayesian algorithm guaranteed to satisfy assumptions

High Level Questions

Can we develop dynamic policies that adapt to their environment over time?

Without making assumptions on how the environment evolves?

- I.e. prior-free or adversarial

With permissive assumptions on agent behavior?

- Ex-ante optimal: Bayesian agents *want to* satisfy assumptions
- Ex-post feasible: non-Bayesian algorithm guaranteed to satisfy assumptions

Framework

Repeated interaction between policymaker and single agent

Hidden state of nature observed after each interaction

Minimum viable case?

Framework

Repeated interaction between policymaker and single agent

Hidden state of nature observed after each interaction

Minimum viable case?

Framework

Repeated interaction between policymaker and single agent

Hidden state of nature observed after each interaction

Minimum viable case?

Contribution

Standard behavioral assumptions are insufficient, allow for odd behaviors

Refine existing assumptions to counterfactual calibration

- Agent fully and consistently exploits any private information

Propose calibrated policy that adapts over time using historical data

Conditions under policymaker's regret is bounded relative to best static policy

Contribution

Standard behavioral assumptions are insufficient, allow for odd behaviors

Refine existing assumptions to counterfactual calibration

- Agent fully and consistently exploits any private information

Propose calibrated policy that adapts over time using historical data

Conditions under policymaker's regret is bounded relative to best static policy

Contribution

Standard behavioral assumptions are insufficient, allow for odd behaviors

Refine existing assumptions to counterfactual calibration

- Agent fully and consistently exploits any private information

Propose calibrated policy that adapts over time using historical data

Conditions under policymaker's regret is bounded relative to best static policy

Contribution

Standard behavioral assumptions are insufficient, allow for odd behaviors

Refine existing assumptions to counterfactual calibration

- Agent fully and consistently exploits any private information

Propose calibrated policy that adapts over time using historical data

Conditions under policymaker's regret is bounded relative to best static policy

Related Literature

Robust dynamic mechanisms, e.g. Chassang 2013, Penta 2015, Madarász and Prat 2016, Mirrokni, Paes Leme, Tang, and Zuo 2020, Carroll 2020, ...

Data-driven auction design, e.g. Blum and Hartline 2005, Elkind 2007, Balcan, Blum, Hartline, and Mansour 2008, Cole and Roughgarden 2014, ...

Regret-based behavioral assumptions, e.g. Foster and Vohra 1997, Nekipelov, Syrgkanis, and Tardos 2015, Braverman, Mao, Schneider, and Weinberg 2018, ...

Statistical learning in incomplete information games, e.g. Liang 2020, Immorlica, Mao, Slivkins, and Wu 2020, Cummings, Devanur, Huang, and Wang 2020, ...

Related Literature

Robust dynamic mechanisms, e.g. Chassang 2013, Penta 2015, Madarász and Prat 2016, Mirrokni, Paes Leme, Tang, and Zuo 2020, Carroll 2020, ...

Data-driven auction design, e.g. Blum and Hartline 2005, Elkind 2007, Balcan, Blum, Hartline, and Mansour 2008, Cole and Roughgarden 2014, ...

Regret-based behavioral assumptions, e.g. Foster and Vohra 1997, Nekipelov, Syrgkanis, and Tardos 2015, Braverman, Mao, Schneider, and Weinberg 2018, ...

Statistical learning in incomplete information games, e.g. Liang 2020, Immorlica, Mao, Slivkins, and Wu 2020, Cummings, Devanur, Huang, and Wang 2020, ...

Related Literature

Robust dynamic mechanisms, e.g. Chassang 2013, Penta 2015, Madarász and Prat 2016, Mirrokni, Paes Leme, Tang, and Zuo 2020, Carroll 2020, ...

Data-driven auction design, e.g. Blum and Hartline 2005, Elkind 2007, Balcan, Blum, Hartline, and Mansour 2008, Cole and Roughgarden 2014, ...

Regret-based behavioral assumptions, e.g. Foster and Vohra 1997, Nekipelov, Syrgkanis, and Tardos 2015, Braverman, Mao, Schneider, and Weinberg 2018, ...

Statistical learning in incomplete information games, e.g. Liang 2020, Immorlica, Mao, Slivkins, and Wu 2020, Cummings, Devanur, Huang, and Wang 2020, ...

Related Literature

Robust dynamic mechanisms, e.g. Chassang 2013, Penta 2015, Madarász and Prat 2016, Mirrokni, Paes Leme, Tang, and Zuo 2020, Carroll 2020, ...

Data-driven auction design, e.g. Blum and Hartline 2005, Elkind 2007, Balcan, Blum, Hartline, and Mansour 2008, Cole and Roughgarden 2014, ...

Regret-based behavioral assumptions, e.g. Foster and Vohra 1997, Nekipelov, Syrgkanis, and Tardos 2015, Braverman, Mao, Schneider, and Weinberg 2018, ...

Statistical learning in incomplete information games, e.g. Liang 2020, Immorlica, Mao, Slivkins, and Wu 2020, Cummings, Devanur, Huang, and Wang 2020, ...

Model

Stage Game

Policymaker sets policy $p \in \mathcal{P}$, sends message $m \in \mathcal{M}$

Agent chooses response $r \in \mathcal{R}$

Hidden state of nature $s \in \mathcal{S}$

Payoffs $u^P(p, r, s)$ and $u^A(p, r, s)$

Assumption: $\mathcal{P}, \mathcal{M}, \mathcal{R}, \mathcal{S}$ finite

Stage Game

Policymaker sets policy $p \in \mathcal{P}$, sends message $m \in \mathcal{M}$

Agent chooses response $r \in \mathcal{R}$

Hidden state of nature $s \in \mathcal{S}$

Payoffs $u^P(p, r, s)$ and $u^A(p, r, s)$

Assumption: $\mathcal{P}, \mathcal{M}, \mathcal{R}, \mathcal{S}$ finite

Stage Game

Policymaker sets policy $p \in \mathcal{P}$, sends message $m \in \mathcal{M}$

Agent chooses response $r \in \mathcal{R}$

Hidden state of nature $s \in \mathcal{S}$

Payoffs $u^P(p, r, s)$ and $u^A(p, r, s)$

Assumption: $\mathcal{P}, \mathcal{M}, \mathcal{R}, \mathcal{S}$ finite

Stage Game

Policymaker sets policy $p \in \mathcal{P}$, sends message $m \in \mathcal{M}$

Agent chooses response $r \in \mathcal{R}$

Hidden state of nature $s \in \mathcal{S}$

Payoffs $u^P(p, r, s)$ and $u^A(p, r, s)$

Assumption: $\mathcal{P}, \mathcal{M}, \mathcal{R}, \mathcal{S}$ finite

Stage Game

Policymaker sets policy $p \in \mathcal{P}$, sends message $m \in \mathcal{M}$

Agent chooses response $r \in \mathcal{R}$

Hidden state of nature $s \in \mathcal{S}$

Payoffs $u^P(p, r, s)$ and $u^A(p, r, s)$

Assumption: $\mathcal{P}, \mathcal{M}, \mathcal{R}, \mathcal{S}$ finite

Repeated Game

Stage game repeated T times

Period t − 1 → **Period t** → **Period t + 1**

⋮

1. Policymaker sets policy p_t and message m_t
2. Agent chooses response r_t
3. State s_t is observed

⋮

Policymaker's mechanism : $\text{history} \to p_t, m_t$

Agent's strategy : $\text{history}, p_t, m_t \to r_t$

Repeated Game

Stage game repeated T times

Period t − 1 → **Period t** → **Period t + 1**

⋮

1. Policymaker sets policy p_t and message m_t
2. Agent chooses response r_t
3. State s_t is observed

⋮

Policymaker's mechanism : $\text{history} \rightarrow p_t, m_t$

Agent's strategy : $\text{history}, p_t, m_t \rightarrow r_t$

Repeated Game

Stage game repeated T times

Period t − 1 → **Period t** → **Period t + 1**

⋮

1. Policymaker sets policy p_t and message m_t
2. Agent chooses response r_t
3. State s_t is observed

⋮

Policymaker's mechanism : $\text{history} \rightarrow p_t, m_t$

Agent's strategy : $\text{history}, p_t, m_t \rightarrow r_t$

Policymaker's Regret

$r_{1:T}$ = agent's responses to actual policies $p_{1:T}$

$r^p_{1:T}$ = agent's responses to fixed policy p

Definition: policymaker's regret compares actual policies with best fixed policy, i.e.

$$\underbrace{\max_p \frac{1}{T}\sum_{t=1}^{T} u^P(p, r^p_t, s_t)}_{\text{utility under best fixed policy } p} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} u^P(p_t, r_t, s_t)}_{\text{utility under actual policies } p_{1:T}}$$

Note: standard no-regret guarantees do not apply since $r^p_t \neq r_t$.

Policymaker's Regret

$r_{1:T}$ = agent's responses to actual policies $p_{1:T}$

$r^{p}_{1:T}$ = agent's responses to fixed policy p

Definition: policymaker's regret compares actual policies with best fixed policy, i.e.

$$\underbrace{\max_{p} \frac{1}{T}\sum_{t=1}^{T} u^P(p, r_t^p, s_t)}_{\text{utility under best fixed policy } p} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} u^P(p_t, r_t, s_t)}_{\text{utility under actual policies } p_{1:T}}$$

Note: standard no-regret guarantees do not apply since $r_t^p \neq r_t$.

Policymaker's Regret

$r_{1:T}$ = agent's responses to actual policies $p_{1:T}$

$r^p_{1:T}$ = agent's responses to fixed policy p

Definition: policymaker's regret compares actual policies with best fixed policy, i.e.

$$\underbrace{\max_p \frac{1}{T}\sum_{t=1}^{T} u^P(p, r^p_t, s_t)}_{\text{utility under best fixed policy } p} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} u^P(p_t, r_t, s_t)}_{\text{utility under actual policies } p_{1:T}}$$

Note: standard no-regret guarantees do not apply since $r^p_t \neq r_t$.

Policymaker's Regret

$r_{1:T}$ = agent's responses to actual policies $p_{1:T}$

$r^{p}_{1:T}$ = agent's responses to fixed policy p

Definition: policymaker's regret compares actual policies with best fixed policy, i.e.

$$\underbrace{\max_{p} \frac{1}{T}\sum_{t=1}^{T} u^P(p, r_t^p, s_t)}_{\text{utility under best fixed policy } p} - \underbrace{\frac{1}{T}\sum_{t=1}^{T} u^P(p_t, r_t, s_t)}_{\text{utility under actual policies } p_{1:T}}$$

Note: standard no-regret guarantees do not apply since $r_t^p \neq r_t$.

Price Regulation

Running Example

Firm produces indivisible good at cost for price to maximize profit.

- Response r : $\text{cost} \to \text{price}$
- State s = firm's cost, buyer's value

Policymaker regulates price to maximize welfare.

- Policy $p = (\text{price floor}, \text{price cap})$

Outcome is $\text{sale} = \mathbf{1}(\text{value} \geq \text{price}) \cdot \mathbf{1}(\text{price floor} \leq \text{price} \leq \text{price cap})$

- Profit $= \text{sale} \cdot (\text{price} - \text{cost})$
- Welfare $= \text{sale} \cdot (\text{value} - \text{cost})$

Price Regulation

Running Example

Firm produces indivisible good at cost for price to maximize profit.

- Response r : $\mathrm{cost} \to \mathrm{price}$
- State s = firm's cost, buyer's value

Policymaker regulates price to maximize welfare.

- Policy $p = (\mathrm{price\ floor}, \mathrm{price\ cap})$

Outcome is $\mathrm{sale} = \mathbf{1}(\mathrm{value} \geq \mathrm{price}) \cdot \mathbf{1}(\mathrm{price\ floor} \leq \mathrm{price} \leq \mathrm{price\ cap})$

- Profit $= \mathrm{sale} \cdot (\mathrm{price} - \mathrm{cost})$
- Welfare $= \mathrm{sale} \cdot (\mathrm{value} - \mathrm{cost})$

Price Regulation

Running Example

Firm produces indivisible good at cost for price to maximize profit.

- Response r : $\text{cost} \to \text{price}$
- State s $=$ firm's cost, buyer's value

Policymaker regulates price to maximize welfare.

- Policy p $= (\text{price floor}, \text{price cap})$

Outcome is $\text{sale} = \mathbf{1}(\text{value} \geq \text{price}) \cdot \mathbf{1}(\text{price floor} \leq \text{price} \leq \text{price cap})$

- Profit $= \text{sale} \cdot (\text{price} - \text{cost})$
- Welfare $= \text{sale} \cdot (\text{value} - \text{cost})$

Repeated Price Regulation

Running Example

Sequence of buyers t with value_t

State $s_t = (\text{value}_t, \text{cost}_t)$ observed after period t

- Replace value_t with sale_t if needed

$\text{price}_t, \text{price floor}_t, \text{price cap}_t$ depend on observed history

Repeated Price Regulation

Running Example

Sequence of buyers t with value_t

State $s_t = (\text{value}_t, \text{cost}_t)$ observed after period t

- Replace value_t with sale_t if needed

$\text{price}_t, \text{price floor}_t, \text{price cap}_t$ depend on observed history

Repeated Price Regulation

Running Example

Sequence of buyers t with value_t

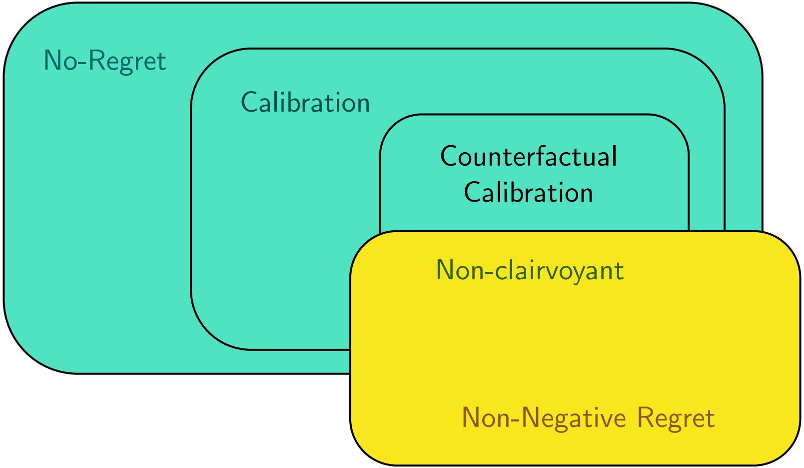
State $s_t = (\mathrm{value}_t, \mathrm{cost}_t)$ observed after period t

- Replace value_t with sale_t if needed

$\mathrm{price}_t, \mathrm{price\ floor}_t, \mathrm{price\ cap}_t$ depend on observed history

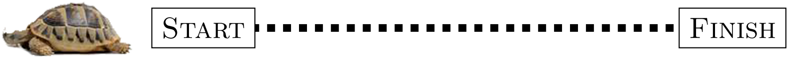
Agent's Behavior

Preview of Definitions



Dealing with Information

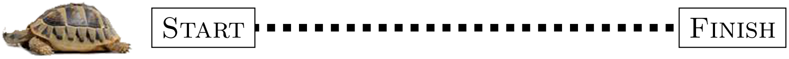
Tortoise travels 1km in 1h : uninformed agent satisfies no-regret.



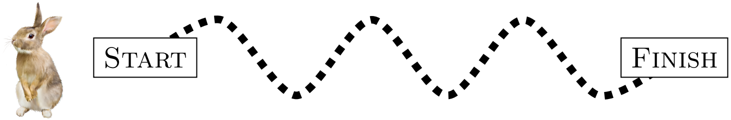
Hare travels 1km in 1h : informed agent satisfies no-regret.

Dealing with Information

Tortoise travels 1km in 1h : uninformed agent satisfies no-regret.



Hare travels 1km in 1h : informed agent satisfies no-regret.



Dealing with Information

Tortoise travels 1km in 1h : uninformed agent satisfies no-regret.

START ········ FINISH

Hare travels 1km in 2min : informed agent satisfies no-regret conditioned on her information.

START ········ FINISH

Agent's Conditional Regret

Definition: agent's regret conditioned on information I_t is

$$\max_{h:\mathcal{I}\to\mathcal{R}} \quad \underbrace{\frac{1}{T}\sum_{t=1}^{T} u^A(p_t, h(I_t), s_t)}_{\text{utility under modified responses } h(I_t)} \quad - \quad \underbrace{\frac{1}{T}\sum_{t=1}^{T} u^A(p_t, r_t, s_t)}_{\text{utility under actual responses } r_{1:T}}$$

where modification rule h maps information I_t to response $h(I_t)$.

No-Regret

Definition: regret = regret conditioned on public information

$$I_t = (p_t, m_t)$$

Definition: no-regret = regret $\to 0$ as $T \to \infty$

No-Regret

Definition: regret = regret conditioned on public information

$$I_t = (p_t, m_t)$$

Definition: no-regret = regret $\to 0$ as $T \to \infty$

Behavior Reveals Information

Game of **rock-paper-scissors** between agent and nature.

$s_t =$	R	P	S	R	P	S	R	P	S	R	P	S

An uninformed strategy; no correlation between response and state

$r_t =$	S	R	R	P	P	P	S	P	R	R	P	R
$u_t^A =$	-1	-1	1	1	0	-1	-1	0	1	0	0	1

Behavior Reveals Information

Game of **rock-paper-scissors** between agent and nature.

$s_t =$ R P S R P S R P S R P S

An uninformed strategy; no correlation between response and state

$r_t =$ S R R P P P S P R R P R

$u_t^A =$ -1 -1 1 1 0 -1 -1 0 1 0 0 1

Behavior Reveals Information

Game of **rock-paper-scissors** between agent and nature.

$s_t =$ R P S R P S R P S R P S

An informed strategy; perfect correlation between response and state

$r_t =$ R P S R P S R P S R P S

$u_t^A =$ 1 1 1 1 1 1 1 1 1 1 1 1

Calibration

Definition: internal regret = regret conditioned on information revealed on-path, i.e.

$$I_t = (p_t, m_t, r_t)$$

Definition: calibration = internal regret $\rightarrow 0$ as $T \rightarrow \infty$

Calibration

Definition: internal regret = regret conditioned on information revealed on-path, i.e.

$$I_t = (p_t, m_t, r_t)$$

Definition: calibration = internal regret $\to 0$ as $T \to \infty$

No-Regret to Calibration

Example: no-regret allows strange behavior that calibration rules out

$s_t =$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$

Tortoise strategy: uninformed, optimal, no-regret

$r_t =$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$

$u_t^A =$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$

No-Regret to Calibration

Example: no-regret allows strange behavior that calibration rules out

$s_t =$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{R}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$

Tortoise strategy: uninformed, optimal, no-regret

$r_t =$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$ $\boxed{P}$

$u_t^A =$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{1}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$ $\boxed{0}$

No-Regret to Calibration

Example: no-regret allows strange behavior that calibration rules out

$s_t =$ R R R R R R P P P P P P

Lazy hare strategy: informed, suboptimal, no-regret

$r_t =$ R R R R R R S S S S S S

$u_t^A =$ 0 0 0 0 0 0 1 1 1 1 1 1

Impossibility Result

Proposition

There exists a strategy for the agent where no mechanism can guarantee non-trivial bound on policymaker's regret across all $s_{1:T}$ where agent's strategy is calibrated.

Takeaway: calibration is not enough for low-regret policy design

- Even if we know the agent's strategy in advance

Impossibility Result

Proposition

There exists a strategy for the agent where no mechanism can guarantee non-trivial bound on policymaker's regret across all $s_{1:T}$ *where agent's strategy is calibrated.*

Takeaway: calibration is not enough for low-regret policy design

- Even if we know the agent's strategy in advance

Counterfactual Calibration

Definition: counterfactual internal regret = regret conditioned on information revealed on- and off-path, i.e.

$$I_t = \left(p_t, m_t, r_t, r_t^{p_1}, \ldots, r_t^{p_n}\right)$$

Definition: counterfactual calibration = CIR $\rightarrow 0$ as $T \rightarrow \infty$

Counterfactual Calibration

Definition: counterfactual internal regret = regret conditioned on information revealed on- and off-path, i.e.

$$I_t = \left(p_t, m_t, r_t, r_t^{p_1}, \ldots, r_t^{p_n}\right)$$

Definition: counterfactual calibration = CIR $\to 0$ as $T \to \infty$

Calibration to Counterfactual Calibration

Example: calibration allows strange behavior that counterfactual calibration rules out

$s_t =$	R	R	R	R	R	R	P	P	P	P	P	P

If policymaker follows mechanism, play tortoise strategy

$r_t =$	P	P	P	P	P	P	P	P	P	P	P	P
$u_t^A =$	1	1	1	1	1	1	0	0	0	0	0	0

If policymaker follows fixed policy, play active hare strategy

$r_t =$	P	P	P	P	P	P	S	S	S	S	S	S
$u_t^A =$	1	1	1	1	1	1	1	1	1	1	1	1

Calibration to Counterfactual Calibration

Example: calibration allows strange behavior that counterfactual calibration rules out

$s_t =$	R	R	R	R	R	R	P	P	P	P	P	P

If policymaker follows mechanism, play tortoise strategy

$r_t =$	P	P	P	P	P	P	P	P	P	P	P	P
$u_t^A =$	1	1	1	1	1	1	0	0	0	0	0	0

If policymaker follows fixed policy, play active hare strategy

$r_t =$	P	P	P	P	P	P	S	S	S	S	S	S
$u_t^A =$	1	1	1	1	1	1	1	1	1	1	1	1

Calibration to Counterfactual Calibration

Example: calibration allows strange behavior that counterfactual calibration rules out

$s_t =$ | R | R | R | R | R | R | P | P | P | P | P | P |

If policymaker follows mechanism, play tortoise strategy

$r_t =$	P	P	P	P	P	P	P	P	P	P	P	P
$u_t^A =$	1	1	1	1	1	1	0	0	0	0	0	0

If policymaker follows fixed policy, play active hare strategy

$r_t =$	P	P	P	P	P	P	S	S	S	S	S	S
$u_t^A =$	1	1	1	1	1	1	1	1	1	1	1	1

Non-clairvoyance

Definition: non-clairvoyance = counterfactual calibration & non-negative regret

Intuition:

1. Counterfactual calibration $\implies$ agent fully exploits her private info
2. Non-negative regret $\implies$ agent doesn't outperform best use of public info
3. Therefore, her private info must not be useful

Non-clairvoyance

Definition: non-clairvoyance = counterfactual calibration & non-negative regret

Intuition:

1. Counterfactual calibration $\implies$ agent fully exploits her private info
2. Non-negative regret $\implies$ agent doesn't outperform best use of public info
3. Therefore, her private info must not be useful

Non-clairvoyance

Definition: non-clairvoyance = counterfactual calibration & non-negative regret

Intuition:

1. Counterfactual calibration $\implies$ agent fully exploits her private info
2. Non-negative regret $\implies$ agent doesn't outperform best use of public info
3. Therefore, her private info must not be useful

Non-clairvoyance

Definition: non-clairvoyance = counterfactual calibration & non-negative regret

Intuition:

1. Counterfactual calibration $\implies$ agent fully exploits her private info
2. Non-negative regret $\implies$ agent doesn't outperform best use of public info
3. Therefore, her private info must not be useful

Calibrated Policy

Preview of Mechanism

In each period t...

1. Form probabilistic forecast of state s_t
2. Assume agent shares forecast
3. Choose ϵ-robust policy based on forecast

Informal result: works well when agent is non-clairvoyant

Preview of Mechanism

In each period t...

1. Form probabilistic forecast of state s_t
2. Assume agent shares forecast
3. Choose ϵ-robust policy based on forecast

Informal result: works well when agent is non-clairvoyant

Preview of Mechanism

In each period t...

1. Form probabilistic forecast of state s_t
2. Assume agent shares forecast
3. Choose ϵ-robust policy based on forecast

Informal result: works well when agent is non-clairvoyant

Preview of Mechanism

In each period t...

1. Form probabilistic forecast of state s_t
2. Assume agent shares forecast
3. Choose ϵ-robust policy based on forecast

Informal result: works well when agent is non-clairvoyant

ϵ-Robustness

Stage Game

π_t^s = state distribution

Definition: mixed response π^r is ϵ-best reply to policy p if

$$\underbrace{\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p, r', s)\right]}_{\text{optimal utility}} - \underbrace{\mathrm{E}_{\pi^s,\pi^r}\left[u^A(p, r, s)\right]}_{\text{actual utility}} \leq \epsilon$$

Definition: policymaker's worst-case utility given ϵ-best reply is

$$\mathrm{WC}_\epsilon(p, \pi^s,) = \min_{\pi^r} \mathrm{E}_{\pi^s,\pi^r}\left[u^P(p, r, s)\right] \quad \text{s.t. } \pi^r \text{ is } \epsilon\text{-best reply}$$

ϵ-Robustness *Stage Game*

π_t^s = state distribution

Definition: mixed response π^r is ϵ-best reply to policy p if

$$\underbrace{\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p, r', s)\right]}_{\text{optimal utility}} - \underbrace{\mathrm{E}_{\pi^s,\pi^r}\left[u^A(p, r, s)\right]}_{\text{actual utility}} \leq \epsilon$$

Definition: policymaker's worst-case utility given ϵ-best reply is

$$\mathrm{WC}_\epsilon(p, \pi^s,) = \min_{\pi^r} \mathrm{E}_{\pi^s,\pi^r}\left[u^P(p, r, s)\right] \quad \text{s.t. } \pi^r \text{ is } \epsilon\text{-best reply}$$

ϵ-Robustness

Stage Game

π_t^s = state distribution

Definition: mixed response π^r is ϵ-best reply to policy p if

$$\underbrace{\max_{r'} \mathrm{E}_{\pi^s}\left[u^A(p, r', s)\right]}_{\text{optimal utility}} - \underbrace{\mathrm{E}_{\pi^s, \pi^r}\left[u^A(p, r, s)\right]}_{\text{actual utility}} \leq \epsilon$$

Definition: policymaker's worst-case utility given ϵ-best reply is

$$\mathrm{WC}_\epsilon(p, \pi^s,) = \min_{\pi^r} \mathrm{E}_{\pi^s, \pi^r}\left[u^P(p, r, s)\right] \quad \text{s.t. } \pi^r \text{ is } \epsilon\text{-best reply}$$

Cost of ϵ-Robustness

Stage Game

Definition: policymaker's best-case utility given ϵ-best reply is

$$\mathrm{BC}_\epsilon(p, \pi^s) = \max_{\pi^r} \mathrm{E}_{\pi^s, \pi^r}\left[u^P(p, r, s)\right] \quad \text{s.t. } \pi^r \text{ is } \epsilon\text{-best reply}$$

Definition: cost of ϵ-robustness is

$$\mathrm{CoR}_\epsilon(p, \pi^s) := \mathrm{BC}_\epsilon(p, \pi^s) - \mathrm{WC}_\epsilon(p, \pi^s)$$

Cost of ϵ-Robustness

Stage Game

Definition: policymaker's best-case utility given ϵ-best reply is

$$\mathrm{BC}_\epsilon(p, \pi^s) = \max_{\pi^r} \mathrm{E}_{\pi^s, \pi^r}\left[u^P(p, r, s)\right] \quad \text{s.t. } \pi^r \text{ is } \epsilon\text{-best reply}$$

Definition: cost of ϵ-robustness is

$$\mathrm{CoR}_\epsilon(p, \pi^s) := \mathrm{BC}_\epsilon(p, \pi^s) - \mathrm{WC}_\epsilon(p, \pi^s)$$

Calibrated Policy

ϵ = robustness parameter

$\tilde{\pi}_t^s$ = calibrated forecast with grid width δ

Policy $\mathbf{p_t}$:= ϵ-robust policy assuming forecast is correct, i.e.

$$p_t \in \arg\max_p \mathrm{WC}_\epsilon(p, \tilde{\pi}_t^s)$$

Message $\mathbf{m_t}$:= forecast $\tilde{\pi}_t^s$

Calibrated Policy

ϵ = robustness parameter

$\tilde{\pi}_t^s$ = calibrated forecast with grid width δ

Policy $\mathbf{p_t}$:= ϵ-robust policy assuming forecast is correct, i.e.

$$p_t \in \arg\max_p \mathrm{WC}_\epsilon(p, \tilde{\pi}_t^s)$$

Message $\mathbf{m_t}$:= forecast $\tilde{\pi}_t^s$

Calibrated Policy

ϵ = robustness parameter

$\tilde{\pi}_t^s$ = calibrated forecast with grid width δ

Policy $\mathbf{p_t}$:= ϵ-robust policy assuming forecast is correct, i.e.

$$p_t \in \arg\max_p \mathrm{WC}_\epsilon(p, \tilde{\pi}_t^s)$$

Message $\mathbf{m_t}$:= forecast $\tilde{\pi}_t^s$

Calibrated Policy

ϵ = robustness parameter

$\tilde{\pi}_t^s$ = calibrated forecast with grid width δ

Policy $\mathbf{p_t}$:= ϵ-robust policy assuming forecast is correct, i.e.

$$p_t \in \arg\max_p \mathrm{WC}_\epsilon(p, \tilde{\pi}_t^s)$$

Message $\mathbf{m_t}$:= forecast $\tilde{\pi}_t^s$

Assumptions

1. Agent is non-clairvoyant
   counterfactual calibration + non-negative regret

2. Information useless to agent under any policy $\Rightarrow$ not harmful to policymaker
   technical assumption on the stage game

Assumptions

1. Agent is non-clairvoyant
   counterfactual calibration + non-negative regret

2. Information useless to agent under any policy $\Rightarrow$ not harmful to policymaker
   technical assumption on the stage game

Regret Bound

Theorem

Policymaker's regret from calibrated policy is less than

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathrm{CoR}_\epsilon(p_t,\tilde{\pi}_t^s)}_{\text{cost of robustness}} + \underbrace{\frac{1}{\epsilon}}_{\text{sensitivity}}\left(\underbrace{O\left(\mathrm{CIR}_T\right)}_{\text{agent miscalibration}} + \underbrace{\tilde{O}\left(\frac{\sqrt{|\mathcal{S}|\cdot N_\delta(\Delta(\mathcal{S}))}}{T^{1/4}}+\sqrt{\delta}\right)}_{\text{forecast miscalibration}}\right)$$

$\tilde{\pi}_t^s$ = *forecast with grid width* δ

CIR_T = *counterfactual internal regret*

$\mathcal{S}$ = *state space*
$\Delta(\mathcal{S})$ = *state distributions*
$N_\delta(\cdot)$ = δ*-covering number*

Tradeoff: $\epsilon\uparrow \implies$ sensitivity to miscalibration $\downarrow$ & cost of robustness $\uparrow$

Regret Bound

Theorem

Policymaker's regret from calibrated policy is less than

$$\underbrace{\frac{1}{T}\sum_{t=1}^{T}\mathrm{CoR}_\epsilon(p_t,\tilde{\pi}_t^s)}_{\text{cost of robustness}} + \underbrace{\frac{1}{\epsilon}}_{\text{sensitivity}}\left(\underbrace{O\left(\mathrm{CIR}_T\right)}_{\text{agent miscalibration}} + \underbrace{\tilde{O}\left(\frac{\sqrt{|\mathcal{S}|\cdot N_\delta(\Delta(\mathcal{S}))}}{T^{1/4}}+\sqrt{\delta}\right)}_{\text{forecast miscalibration}}\right)$$

$\tilde{\pi}_t^s =$ *forecast with grid width* δ

$\mathrm{CIR}_T =$ *counterfactual internal regret*

$\mathcal{S} =$ *state space*
$\Delta(\mathcal{S}) =$ *state distributions*
$N_\delta(\cdot) =$ δ*-covering number*

Tradeoff: $\epsilon\uparrow \implies$ sensitivity to miscalibration $\downarrow$ & cost of robustness $\uparrow$

Robustness Lemma

Lemma

For any distribution π^s, *policy* p, *and constants* $\epsilon' > \epsilon > 0$,

$$\mathrm{WC}_{\epsilon'}(p, \pi^s) \geq \mathrm{WC}_{\epsilon}(p, \pi^s) - O\left(\frac{\epsilon' - \epsilon}{\epsilon}\right)$$

Calibrated Price Regulation

Running Example

Typical tradeoff:

- price cap_t too large $\implies$ price_t too large, fewer sales
- price cap_t too small $\implies$ risk of $\text{cost}_t > \text{price cap}_t$, firm shutdown

How to balance tradeoff depends on market conditions

- Predict market conditions = forecast of $(\text{value}_t, \text{cost}_t)$
- Even if $\text{value}_{1:t-1}$ not observed, feasible if $(\text{sale}_{1:t-1}, \text{cost}_{1:t-1})$ observed

Calibrated policy assumes forecast true and optimizes in stage game.

- Firm's beliefs $\approx$ forecast $\implies$ ϵ-optimal pricing w.r.t. forecast

Calibrated Price Regulation

Running Example

Typical tradeoff:

- price cap_t too large $\implies$ price_t too large, fewer sales
- price cap_t too small $\implies$ risk of $\text{cost}_t > \text{price cap}_t$, firm shutdown

How to balance tradeoff depends on market conditions

- Predict market conditions = forecast of $(\text{value}_t, \text{cost}_t)$
- Even if $\text{value}_{1:t-1}$ not observed, feasible if $(\text{sale}_{1:t-1}, \text{cost}_{1:t-1})$ observed

Calibrated policy assumes forecast true and optimizes in stage game.

- Firm's beliefs $\approx$ forecast $\implies$ ϵ-optimal pricing w.r.t. forecast

Calibrated Price Regulation

Running Example

Typical tradeoff:

- price cap_t too large $\implies$ price_t too large, fewer sales
- price cap_t too small $\implies$ risk of $\text{cost}_t > \text{price cap}_t$, firm shutdown

How to balance tradeoff depends on market conditions

- Predict market conditions = forecast of $(\text{value}_t, \text{cost}_t)$
- Even if $\text{value}_{1:t-1}$ not observed, feasible if $(\text{sale}_{1:t-1}, \text{cost}_{1:t-1})$ observed

Calibrated policy assumes forecast true and optimizes in stage game.

- Firm's beliefs ≈ forecast $\implies$ ϵ-optimal pricing w.r.t. forecast

Calibrated Price Regulation

Running Example

Typical tradeoff:

- $\mathrm{price\,cap}_t$ too large $\implies$ price_t too large, fewer sales
- $\mathrm{price\,cap}_t$ too small $\implies$ risk of $\mathrm{cost}_t > \mathrm{price\,cap}_t$, firm shutdown

How to balance tradeoff depends on market conditions

- Predict market conditions = forecast of $(\mathrm{value}_t, \mathrm{cost}_t)$
- Even if $\mathrm{value}_{1:t-1}$ not observed, feasible if $(\mathrm{sale}_{1:t-1}, \mathrm{cost}_{1:t-1})$ observed

Calibrated policy assumes forecast true and optimizes in stage game.

- Firm's beliefs $\approx$ forecast $\implies$ ϵ-optimal pricing w.r.t. forecast

Conclusion

Ways to Deal with Private Information

1. Assume it doesn't exist
2. Assume it exists but is well-understood
3. Optimize against worst-case private information (extension)
4. Adapt to private information over time (work-in-progress)

Ways to Deal with Private Information

1. Assume it doesn't exist
2. Assume it exists but is well-understood
3. Optimize against worst-case private information (extension)
4. Adapt to private information over time (work-in-progress)

Ways to Deal with Private Information

1. Assume it doesn't exist
2. Assume it exists but is well-understood
3. Optimize against worst-case private information (extension)
4. Adapt to private information over time (work-in-progress)

Ways to Deal with Private Information

1. Assume it doesn't exist
2. Assume it exists but is well-understood
3. Optimize against worst-case private information (extension)
4. Adapt to private information over time (work-in-progress)

Future Directions

Beyond our minimum viable case...

- What if feedback is imperfect?
- Can we incorporate dynamic incentives?
- Multiple agents?

Potential areas of application?

- Price regulation that adapts to changing costs and demand
- Minimum wages that adapt to changing labor markets
- Worker incentives that adapt to changing workplace

Future Directions

Beyond our minimum viable case...

- What if feedback is imperfect?
- Can we incorporate dynamic incentives?
- Multiple agents?

Potential areas of application?

- Price regulation that adapts to changing costs and demand
- Minimum wages that adapt to changing labor markets
- Worker incentives that adapt to changing workplace

Thank you!