“Calibeating”: Beating Forecasters at Their Own Game

Sergiu Hart

June 2024

"Calibeating": Beating Forecasters at Their Own Game

Sergiu Hart

Center for the Study of Rationality
Dept of Mathematics Dept of Economics
The Hebrew University of Jerusalem

hart@huji.ac.il

http://www.ma.huji.ac.il/hart

Joint work with

Dean P. Foster

University of Pennsylvania &

Amazon Research NY

Papers

Papers

- Sergiu Hart
  "Calibration: The Minimax Proof", 1995 [2021]
  www.ma.huji.ac.il/hart/publ.html#calib-minmax

Papers

- Sergiu Hart
  "Calibration: The Minimax Proof", 1995 [2021]
  www.ma.huji.ac.il/hart/publ.html#calib-minmax

- Dean P. Foster and Sergiu Hart
  "Smooth Calibration, Leaky Forecasts, Finite Recall, and Nash Dynamics"
  Games and Economic Behavior 2018
  www.ma.huji.ac.il/hart/publ.html#calib-eq

Papers

Papers

- Dean P. Foster and Sergiu Hart
  "Forecast Hedging and Calibration"
  Journal of Political Economy 2021

  www.ma.huji.ac.il/hart/publ.html#calib-int

Papers

- Dean P. Foster and Sergiu Hart
  “Forecast Hedging and Calibration”
  Journal of Political Economy 2021

  www.ma.huji.ac.il/hart/publ.html#calib-int

- Dean P. Foster and Sergiu Hart
  “ ‘Calibeating’: Beating Forecasters at Their Own Game”
  Theoretical Economics 2023

  www.ma.huji.ac.il/hart/publ.html#calib-beat

Calibration

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"
- Forecaster is **CALIBRATED** if

Calibration

- Forecaster says: "***The probability of rain tomorrow is*** p"
- Forecaster is **CALIBRATED** if
  - for every forecast p:
    in the days when the forecast was p, the proportion of rainy days equals p

Calibration

- Forecaster says: “***The probability of rain tomorrow is*** p”
- Forecaster is **CALIBRATED** if
  - for every forecast p:
    in the days when the forecast was p, the proportion of rainy days equals p
    (or: is close to p in the long run)

Calibration

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

Calibration

CALIBRATION *can be guaranteed*

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
- Hart and Mas-Colell 1996 [publ 2000]: procedure by Blackwell's Approachability

Calibration

CALIBRATION ***can be guaranteed***

(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
- Hart and Mas-Colell 1996 [publ 2000]: procedure by Blackwell's Approachability
- Foster 1999: simple procedure

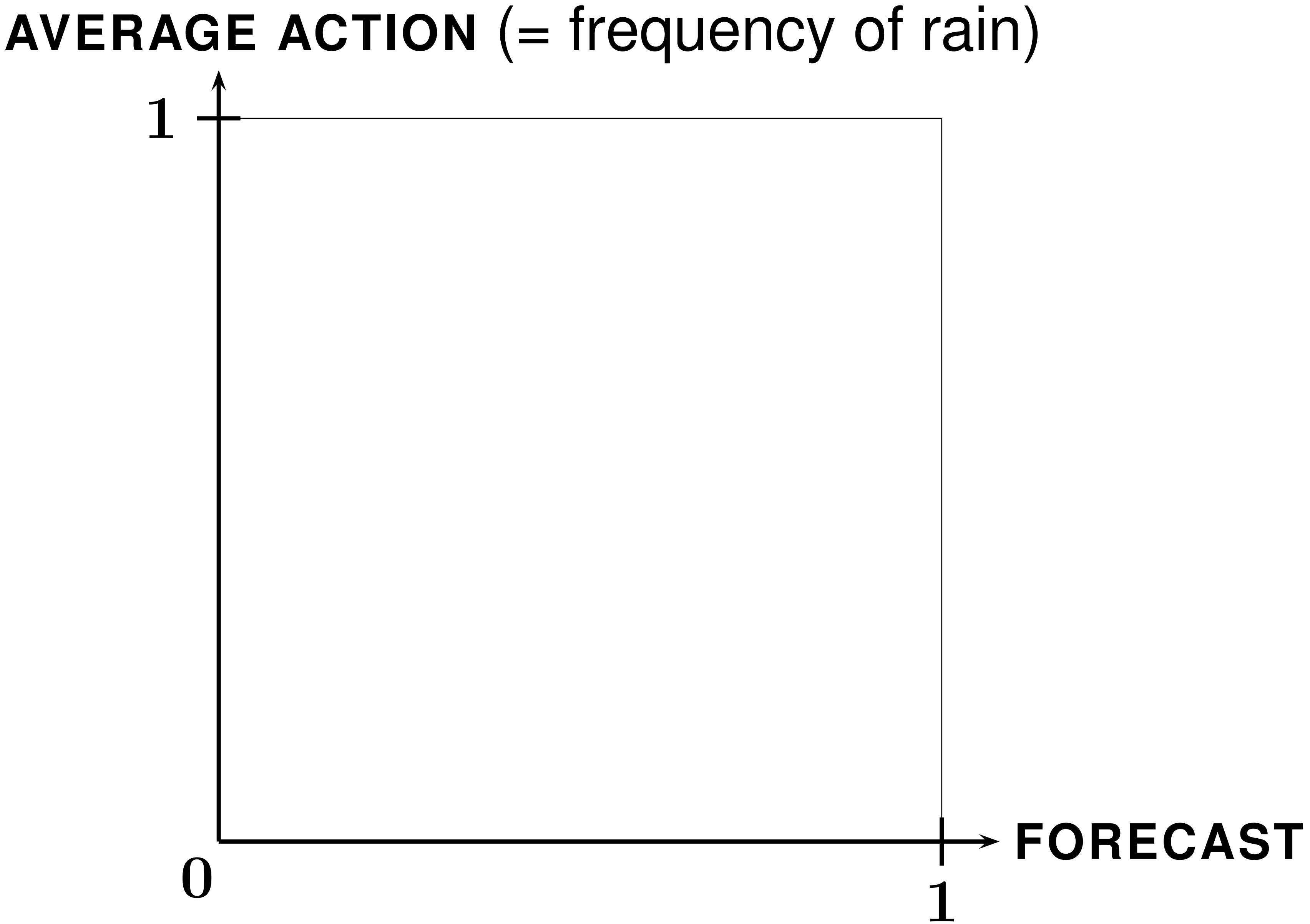
Calibration

CALIBRATION *can be guaranteed*
(no matter what the weather will be)

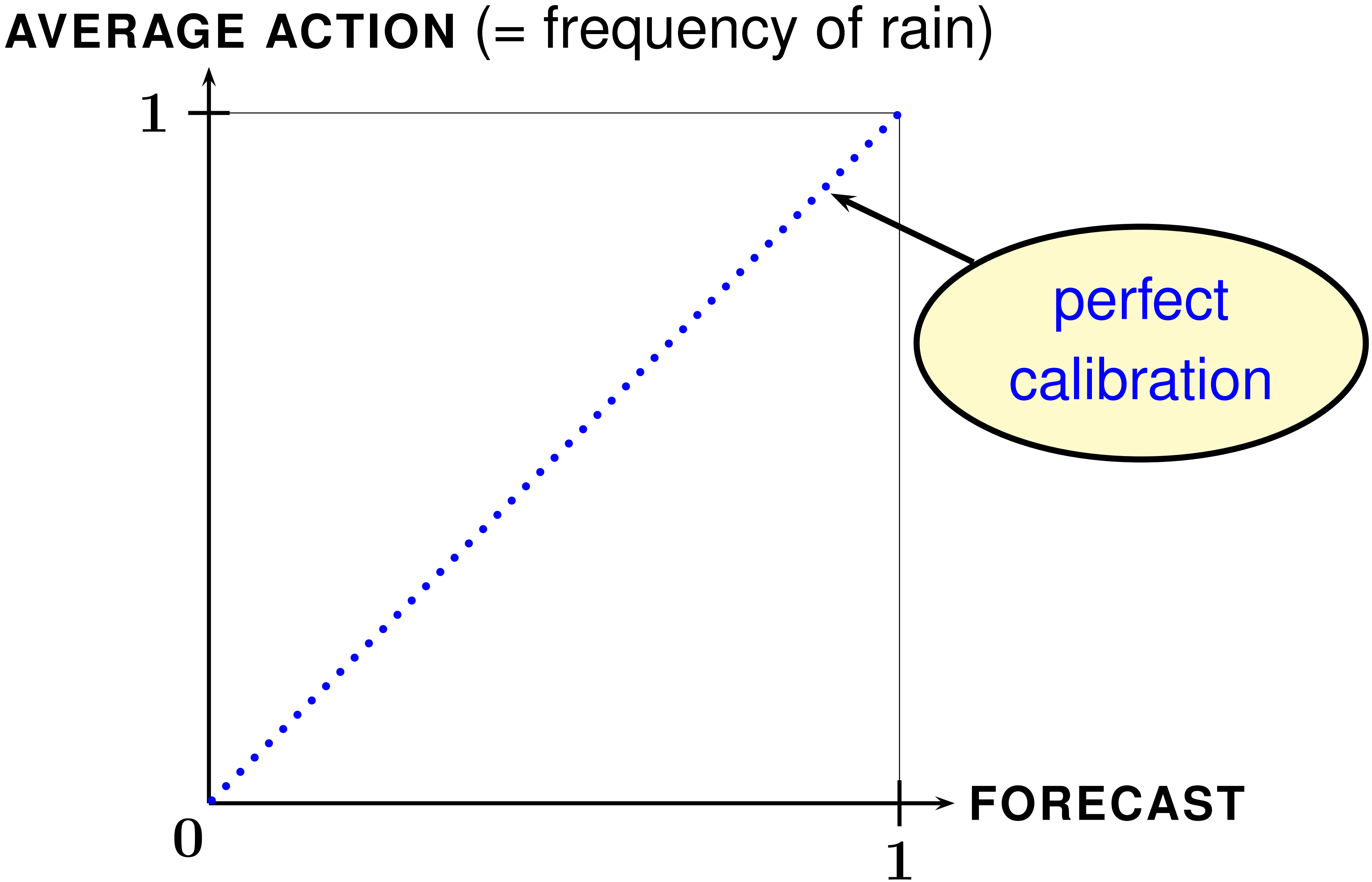
- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
- Hart and Mas-Colell 1996 [publ 2000]: procedure by Blackwell's Approachability
- Foster 1999: simple procedure
- Foster and Hart 2016 [publ 2021]: simplest procedure, by "Forecast Hedging"

Forecast-Hedging

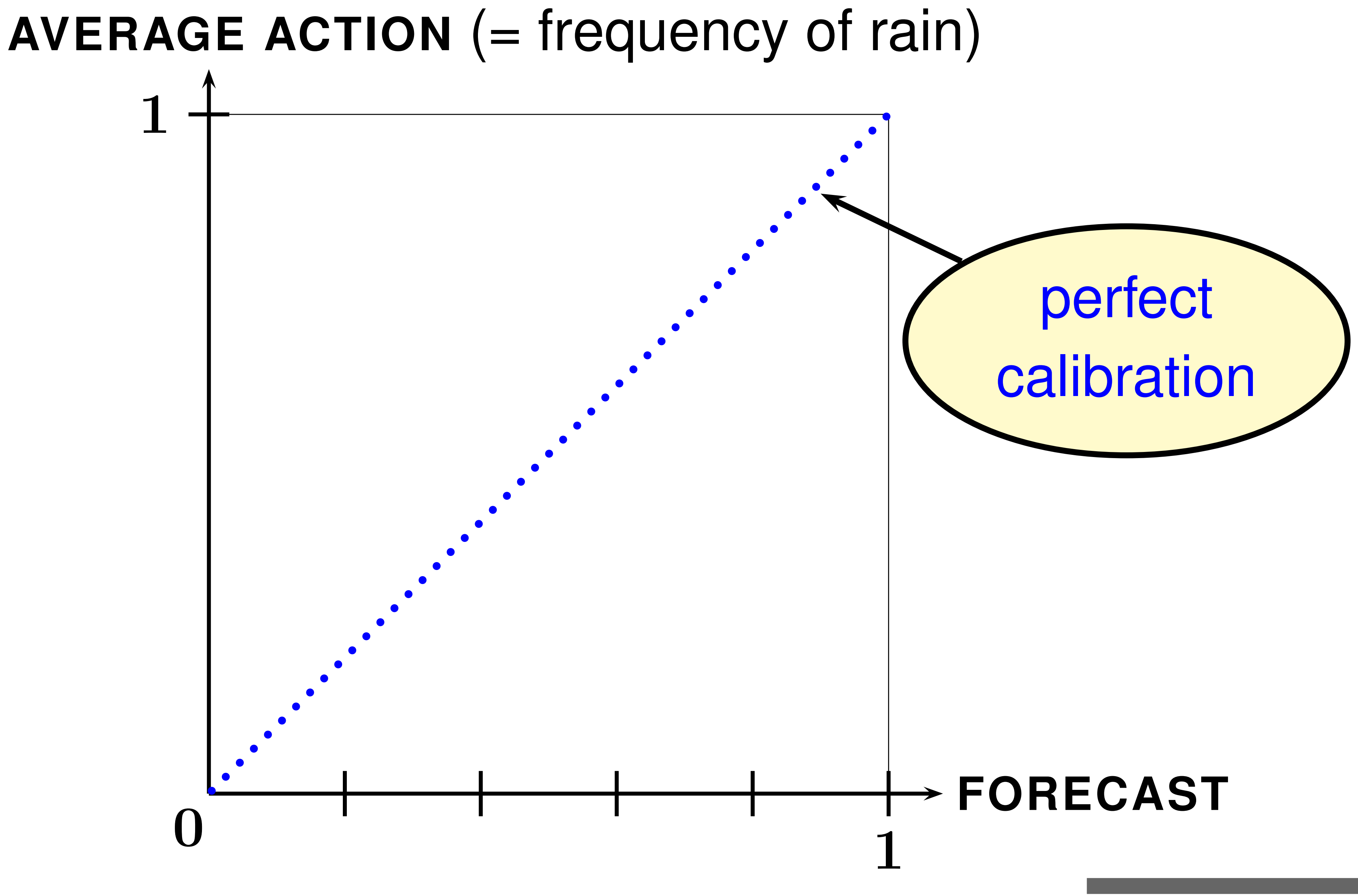
Forecast-Hedging



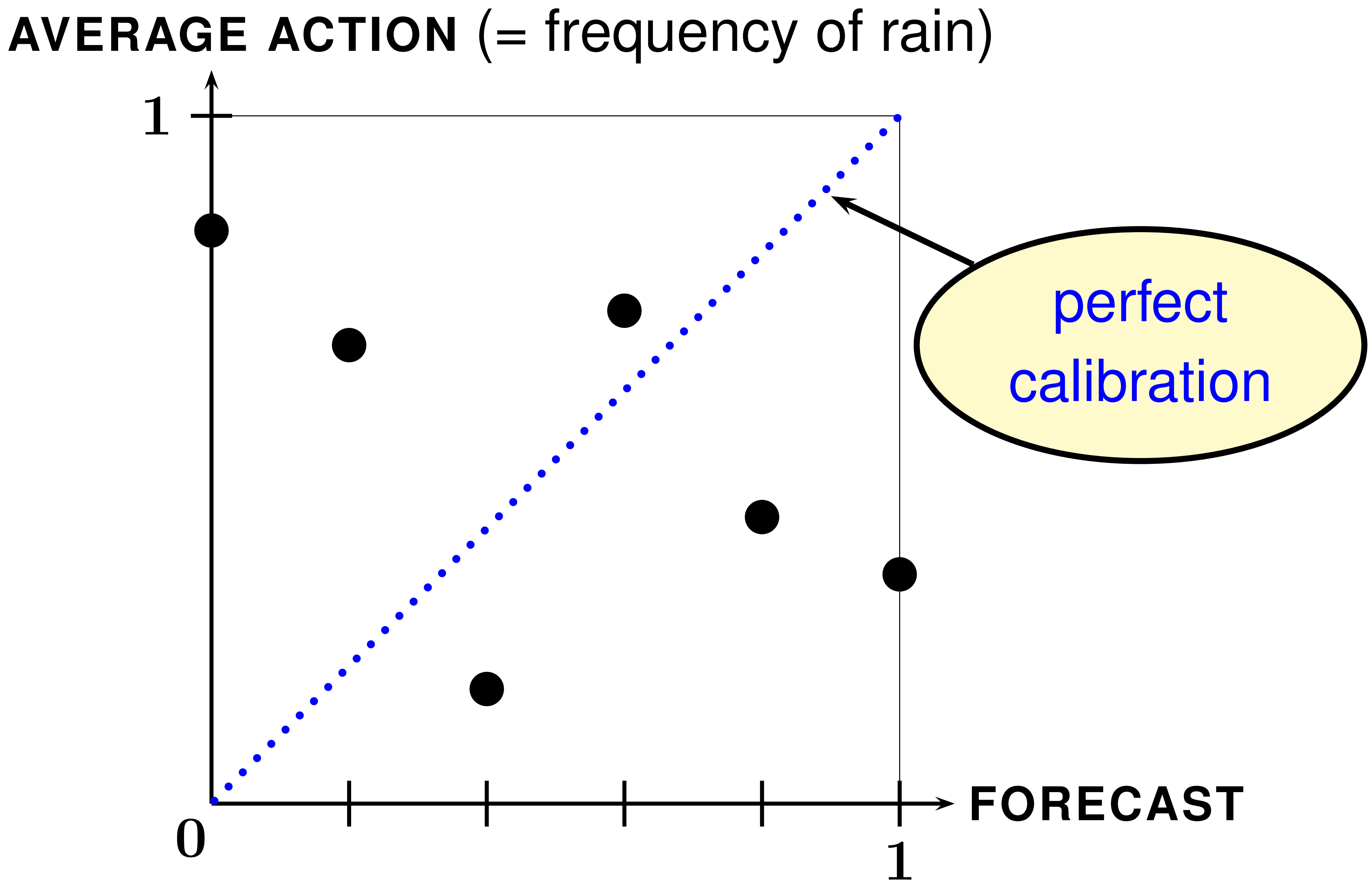
Forecast-Hedging



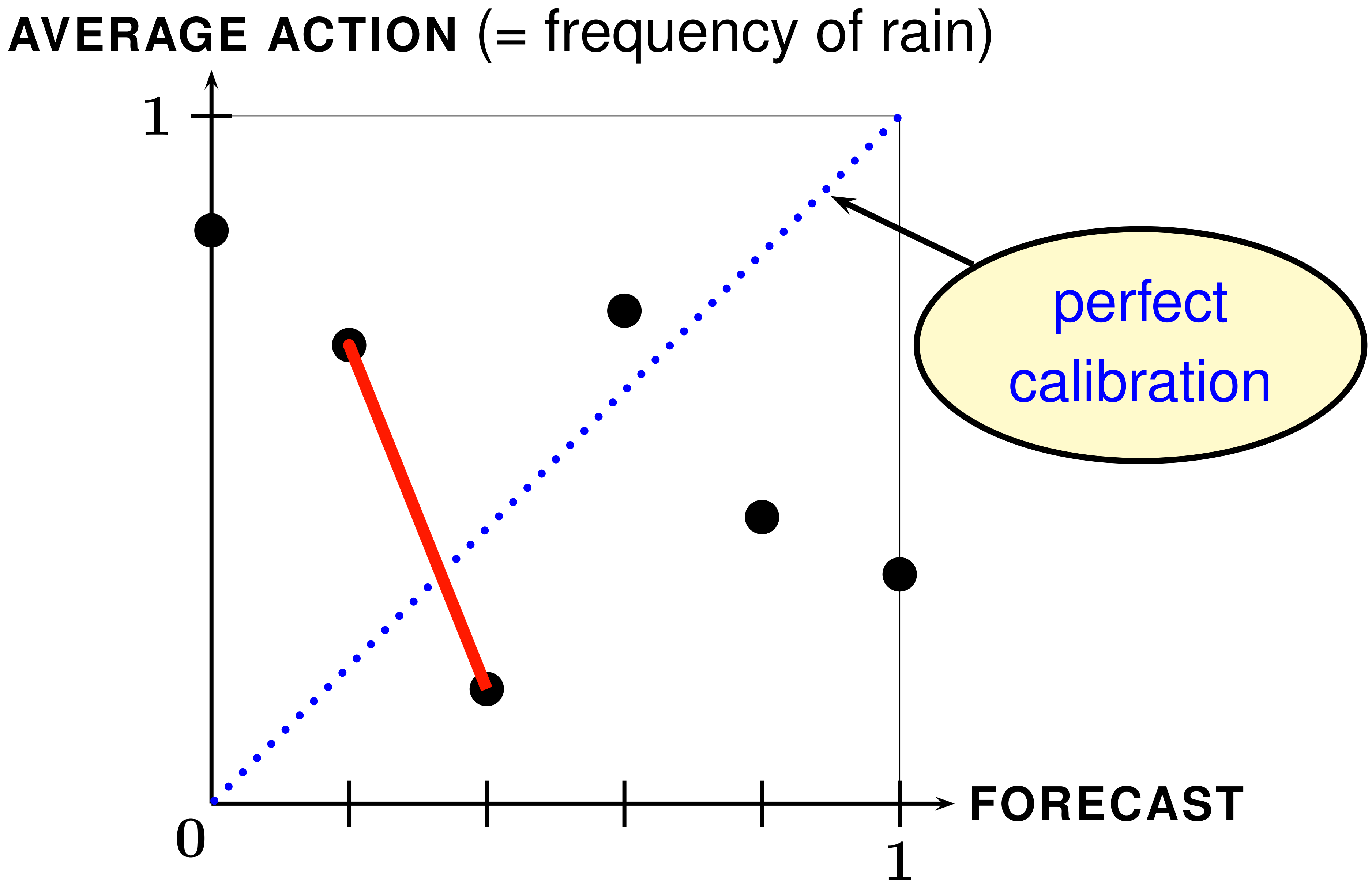
Forecast-Hedging



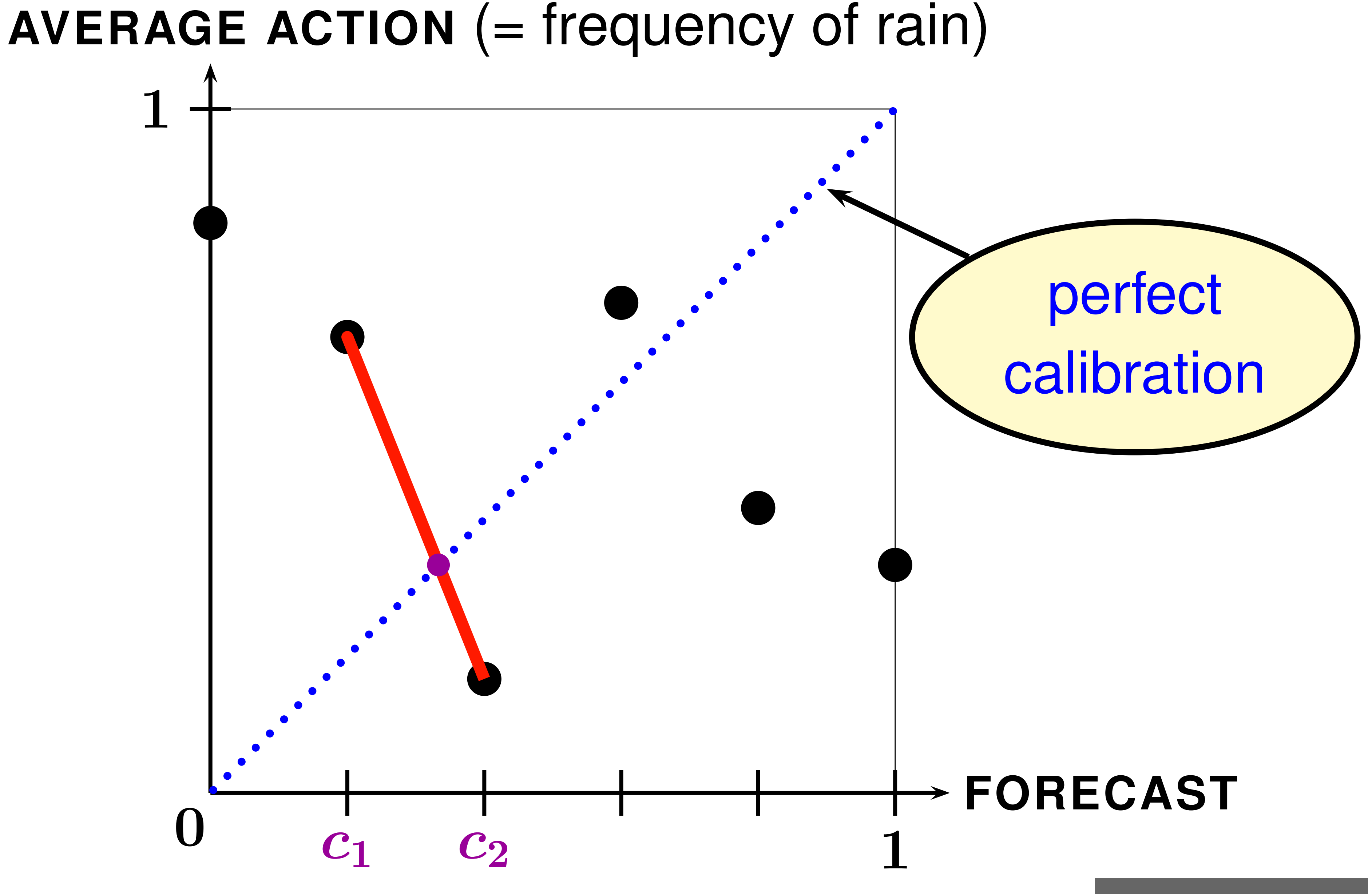
Forecast-Hedging



Forecast-Hedging

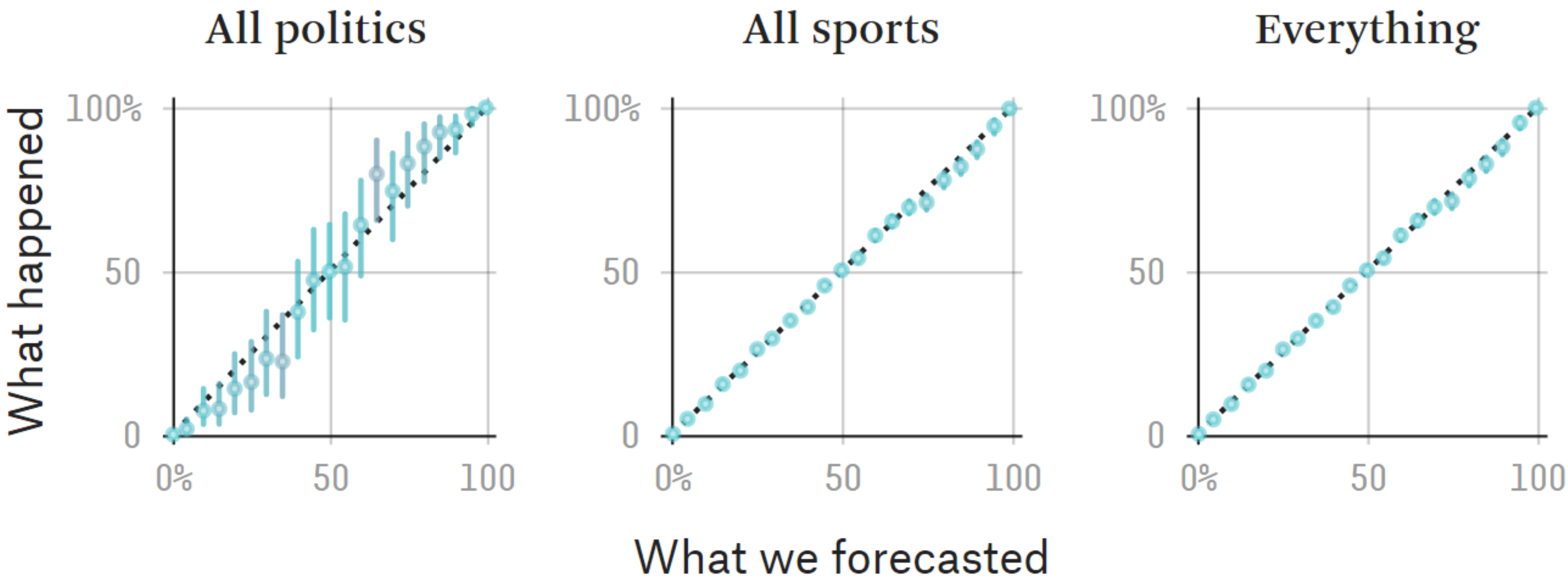


Forecast-Hedging



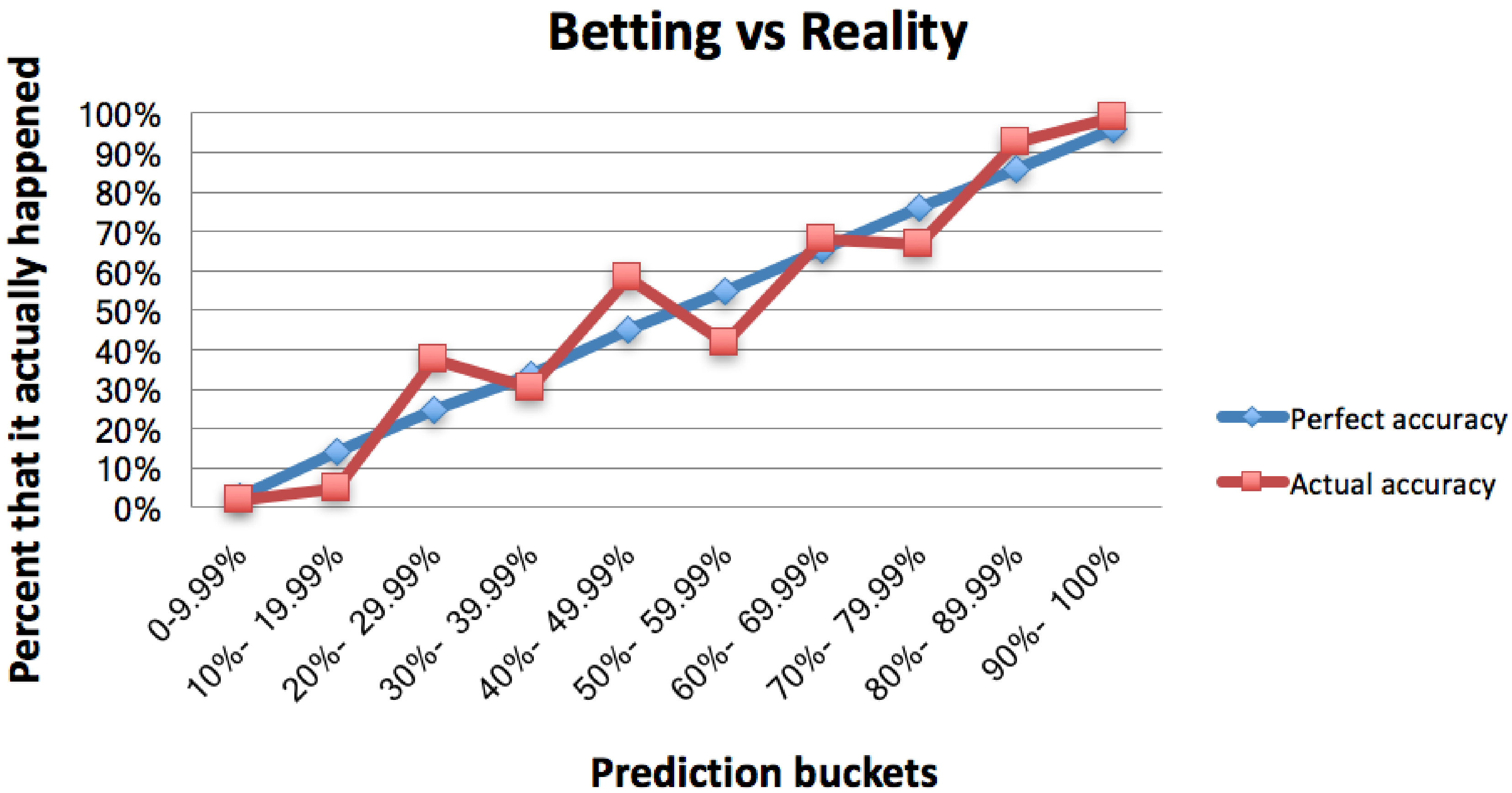
Calibration in Practice

Calibration in Practice



Calibration plots of FiveThirtyEight.com
(as of June 2019)

Calibration in Practice



Calibration plot of ElectionBettingOdds.com
(2016 – 2018)

Example

Example

time	1	2	3	4	5	6	...

Example

time	1	2	3	4	5	6	...
rain	1	0	1	0	1	0	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	

Example

time	1	2	3	4	5	6	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= \mathbf{0}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= \mathbf{0}$

F2: **CALIBRATION** $= \mathbf{0}$

Example

time	**1**	**2**	**3**	**4**	**5**	**6**	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= 0$ **IN-BIN VARIANCE** $= 0$

F2: **CALIBRATION** $= 0$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= \mathbf{0}$ **IN-BIN VARIANCE** $= \mathbf{0}$

F2: **CALIBRATION** $= \mathbf{0}$ **IN-BIN VARIANCE** $= \frac{\mathbf{1}}{\mathbf{4}}$

Notations

Notations

- a_t = action at time t

Notations

- a_t = action at time t
- c_t = forecast at time t

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x

$$\bar{a}(x) = \frac{\sum_{t=1}^{T} \mathbf{1}_x(c_t)\, a_t}{\sum_{t=1}^{T} \mathbf{1}_x(c_t)}$$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$

$$\mathcal{K} = \frac{1}{T} \sum_{t=1}^{T} \| c_t - \bar{a}(c_t) \|^2$$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$
- $\mathcal{R} \equiv \mathcal{R}_T$ = **REFINEMENT** score = average variance inside the bins

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$
- $\mathcal{R} \equiv \mathcal{R}_T$ = **REFINEMENT** score = average variance inside the bins

$$\mathcal{R} = \frac{1}{T} \sum_{t=1}^{T} \|a_t - \bar{a}(c_t)\|^2$$

Notations

- a_t = action at time t
- c_t = forecast at time t
- $\bar{a}(x) \equiv \bar{a}_T(x)$ = average of the actions in all periods where the forecast was x
- $\mathcal{K} \equiv \mathcal{K}_T$ = **CALIBRATION** score = average distance between c_t and $\bar{a}(c_t)$
- $\mathcal{R} \equiv \mathcal{R}_T$ = **REFINEMENT** score = average variance inside the bins

Brier score

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **Brier** (1950) score $=$ average distance between a_t and c_t

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **Brier** (1950) score = average distance between a_t and c_t

$$\mathcal{B} = \frac{1}{T} \sum_{t=1}^{T} \|a_t - c_t\|^2$$

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **Brier** (1950) score $=$ average distance between a_t and c_t

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **BRIER** (1950) score $=$ average distance between a_t and c_t
-

$$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ BRIER (1950) score $=$ average distance between a_t and c_t
- $$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

  BRIER $=$ REFINEMENT $+$ CALIBRATION

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ **Brier** (1950) score $=$ average distance between a_t and c_t
- $$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

  Brier $=$ **REFINEMENT** $+$ **CALIBRATION**

Proof.

$$\mathbb{E}[(X - c)^2] = \mathbb{V}ar(X) + (\bar{X} - c)^2$$

where c is a constant and X is a random variable with $\bar{X} = \mathbb{E}[X]$

Brier score

- $\mathcal{B} \equiv \mathcal{B}_T =$ BRIER (1950) score $=$ average distance between a_t and c_t
- $$\mathcal{B} = \mathcal{R} + \mathcal{K}$$

  BRIER $=$ REFINEMENT $+$ CALIBRATION

Example

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: **CALIBRATION** $= 0$ **IN-BIN VARIANCE** $= 0$

F2: **CALIBRATION** $= 0$ **IN-BIN VARIANCE** $= \frac{1}{4}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: $\mathcal{K} = 0 \quad \mathcal{R} = 0$

F2: $\mathcal{K} = 0 \quad \mathcal{R} = \frac{1}{4}$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
F1	100%	0%	100%	0%	100%	0%	
F2	50%	50%	50%	50%	50%	50%	

F1: $\mathcal{K} = 0 \quad \mathcal{R} = 0 \quad \mathcal{B} = 0$

F2: $\mathcal{K} = 0 \quad \mathcal{R} = \frac{1}{4} \quad \mathcal{B} = \frac{1}{4}$

“Experts”

“Experts”

Testing experts:

“Experts”

Testing experts:

✓ BRIER score

“Experts”

Testing experts:

✓ BRIER score

✗ CALIBRATION score

“Expertise”

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data

"Expertise"

- Recognize ***patterns*** and ***regularities*** in the data
- ***Sort*** the days into bins that consist of similar days

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data
- ***Sort*** the days into bins that consist of similar days
- Make the binning as ***refined*** as possible

“Expertise”

- Recognize ***patterns*** and ***regularities*** in the data
- ***Sort*** the days into bins that consist of similar days
- Make the binning as ***refined*** as possible

$\Leftrightarrow$ ***LOW*** **REFINEMENT SCORE**

“Expertise” and Calibration

“Expertise” and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run

“Expertise” and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever “**EXPERTISE**” one has

“Expertise” and Calibration

- CALIBRATION ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: CALIBRATION procedures ignore whatever “EXPERTISE” one has

Question:

Can one **GAIN CALIBRATION** **without** **LOSING “EXPERTISE”** ?

"Expertise" and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever "**EXPERTISE**" one has

Question:

Can one **GAIN CALIBRATION**

without **LOSING "EXPERTISE"** ?

- Can one get $\mathcal{K}$ to 0 without increasing $\mathcal{R}$?

“Expertise” and Calibration

- **CALIBRATION** ($\mathcal{K} \approx 0$) can always be ***guaranteed*** in the long run
- But: **CALIBRATION** procedures ignore whatever “**EXPERTISE**” one has

Question:

Can one **GAIN CALIBRATION** **without** **LOSING “EXPERTISE”** ?

- Can one get $\mathcal{K}$ to 0 without increasing $\mathcal{R}$?
- Can one decrease $\mathcal{B} = \mathcal{R} + \mathcal{K}$ by $\mathcal{K}$?

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

"Expertise" and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0$$

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R}$$

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R} \qquad \mathcal{B}' = \mathcal{B} - \mathcal{K}$$

“Expertise” and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R} \qquad \mathcal{B}' = \mathcal{B} - \mathcal{K}$$

  - IN RETROSPECT / OFFLINE (when the $\bar{a}(c)$ are known)

"Expertise" and Calibration

- Can one decrease $\mathcal{B}$ by $\mathcal{K}$?
- **Yes:** Replace each forecast c with the corresponding bin average $\bar{a}(c)$

$$\Rightarrow \quad \mathcal{K}' = 0 \qquad \mathcal{R}' = \mathcal{R} \qquad \mathcal{B}' = \mathcal{B} - \mathcal{K}$$

  - **IN RETROSPECT** / **OFFLINE** (when the $\bar{a}(c)$ are known)

Question:

Can one do this **ONLINE** ?

- Consider a forecasting sequence b_t (in a [finite] set B)

- Consider a forecasting sequence $\boldsymbol{b}_t$ (in a [finite] set $\boldsymbol{B}$)
- At each time $\boldsymbol{t}$ generate a forecast $\boldsymbol{c}_t$

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - **ONLINE**: use only b_t and history

- Consider a forecasting sequence $\boldsymbol{b}_t$ (in a [finite] set B)
- At each time t generate a forecast $\boldsymbol{c}_t$
  - **ONLINE**: use only $\boldsymbol{b}_t$ and history
  - such that

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - **ONLINE**: use only b_t and history
  - such that

$$\mathcal{B}_T^{\mathrm{c}} \leq \mathcal{B}_T^{\mathrm{b}} - \mathcal{K}_T^{\mathrm{b}} + \mathrm{o}(1) \quad \text{as } T \to \infty$$

for **ALL** sequences a_t and b_t (uniformly)

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}} = \mathcal{R}^{\mathrm{b}}$$

"Calibeating"

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - **ONLINE**: use only b_t and history
  - such that

$$\boxed{\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}} = \mathcal{R}^{\mathrm{b}}$$

c "**BEATS**" b by b's **CALIBRATION** score

"Calibeating"

- Consider a forecasting sequence b_t (in a [finite] set B)
- At each time t generate a forecast c_t
  - ONLINE: use only b_t and history
  - such that

$$\boxed{\mathcal{B}^{\mathrm{c}} \le \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}} = \mathcal{R}^{\mathrm{b}}$$

c "BEATS" b by b's CALIBRATION score

- GUARANTEED for ALL sequences of actions and forecasts

Example

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	

$\boldsymbol{b}$: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\mathcal{R}^{\mathrm{b}} = 0$ $\mathcal{B}^{\mathrm{b}} = 0.1$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	
$\boldsymbol{c}$	100%	0%	100%	0%	100%	0%	

$\boldsymbol{b}$: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\mathcal{R}^{\mathrm{b}} = 0$ $\mathcal{B}^{\mathrm{b}} = 0.1$

Example

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	
$\boldsymbol{c}$	100%	0%	100%	0%	100%	0%	

$\boldsymbol{b}$: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\mathcal{R}^{\mathrm{b}} = 0$ $\mathcal{B}^{\mathrm{b}} = 0.1$

$\boldsymbol{c}$: $\mathcal{K}^{\mathrm{c}} = 0$ $\mathcal{R}^{\mathrm{c}} = 0$ $\mathcal{B}^{\mathrm{c}} = 0$

Calibeating

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
$\boldsymbol{b}$	80%	40%	80%	40%	80%	40%	
$\boldsymbol{c}$	100%	0%	100%	0%	100%	0%	

$\boldsymbol{b}$: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\mathcal{R}^{\mathrm{b}} = 0$ $\mathcal{B}^{\mathrm{b}} = 0.1$

$\boldsymbol{c}$: $\mathcal{K}^{\mathrm{c}} = 0$ $\mathcal{R}^{\mathrm{c}} = 0$ $\mathcal{B}^{\mathrm{c}} = 0$

c ***calibeats*** b: $\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$

Calibeating

time	$\mathbf{1}$	$\mathbf{2}$	$\mathbf{3}$	$\mathbf{4}$	$\mathbf{5}$	$\mathbf{6}$	...
rain	1	0	1	0	1	0	
b	80%	40%	80%	40%	80%	40%	
c	100%	0%	100%	0%	100%	0%	

b: $\mathcal{K}^{\mathrm{b}} = 0.1$ $\quad \mathcal{R}^{\mathrm{b}} = 0$ $\quad \mathcal{B}^{\mathrm{b}} = 0.1$

c: $\mathcal{K}^{\mathrm{c}} = 0$ $\quad \mathcal{R}^{\mathrm{c}} = 0$ $\quad \mathcal{B}^{\mathrm{c}} = 0$

c ***calibeats*** b: $\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}} = \mathcal{R}^{\mathrm{b}}$

Calibeating

Calibeating

(that was easy ...)

Calibeating

(that was easy ...)

Can one **CALIBEAT** ***in general, non-stationary, situations ?***

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)
- **Bin averages** do not converge

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)
- **Bin averages** do not converge
- **ONLINE**

Calibeating

(that was easy ...)

Can one* CALIBEAT *in general, non-stationary, situations ?

- **Weather** is arbitrary and not stationary
- **Forecasts of** b are arbitrary
- **Binning of** b is not perfect ($\mathcal{R}^{\mathrm{b}} > 0$)
- **Bin averages** do not converge
- **ONLINE**
- **GUARANTEED** (even against adversary)

Calibeating

Calibeating

Theorem

There exists a **CALIBEATING** procedure

A Way to Calibeat

A Way to Calibeat

Theorem

The procedure

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING

A Simple Way to Calibeat

Theorem

The procedure

$$\boldsymbol{c}_t = \bar{\boldsymbol{a}}^{\mathrm{b}}_{t-1}(\boldsymbol{b}_t)$$

GUARANTEES b-**CALIBEATING**

Forecast the average action of the current b-forecast

Proof

Proof

$$\mathbb{V}\mathbf{ar} = \frac{1}{T}\sum_{t=1}^{T}\left\|\boldsymbol{x}_t - \bar{\boldsymbol{x}}_T\right\|^2$$

Proof

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \frac{1}{T}\sum_{t=1}^{T}\left\|x_t - \bar{x}_T\right\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(1-\frac{1}{t}\right)\left\|x_t - \bar{x}_{t-1}\right\|^2
\end{aligned}
$$

Proof

$$\mathbb{V}\text{ar} = \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{\color{red}T}\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{\color{red}t-1}\|^2$$

Proof

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \frac{1}{T}\sum_{t=1}^{T}\left\|x_t - \bar{x}_T\right\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(1 - \frac{1}{t}\right)\left\|x_t - \bar{x}_{t-1}\right\|^2
\end{aligned}
$$

Proof

$$\mathbb{V}\mathrm{ar} = \frac{1}{T}\sum_{t=1}^{T}\|x_t - \bar{x}_T\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T}\left(1 - \frac{1}{t}\right)\|x_t - \bar{x}_{t-1}\|^2$$

$$= \frac{1}{T}\sum_{t=1}^{T}\|x_t - \bar{x}_{t-1}\|^2 - \mathrm{o}(1)$$

Proof

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_T\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{t-1}\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{t-1}\|^2 - \mathrm{o}(1)
\end{aligned}
$$

(*) $\mathrm{o}(1) = \mathrm{O}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{1}{t}\right) = \mathrm{O}\left(\frac{\log T}{T}\right)$

Proof

$$\begin{aligned}\mathbb{V}\text{ar} &= \frac{1}{T}\sum_{t=1}^{T}\|x_t - \bar{x}_T\|^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(1-\frac{1}{t}\right)\|x_t - \bar{x}_{t-1}\|^2 \\ &= \frac{1}{T}\sum_{t=1}^{T}\|x_t - \bar{x}_{t-1}\|^2 - \mathrm{o}(1)\end{aligned}$$

Proof: "Online Variance"

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_T\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \|x_t - \bar{x}_{t-1}\|^2 \\
&= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \|x_t - \bar{x}_{t-1}\|^2}_{\widetilde{\mathbb{V}\mathrm{ar}}} - \mathrm{o}(1) \\
&= \widetilde{\mathbb{V}\mathrm{ar}} - \mathrm{o}(1)
\end{aligned}
$$

Proof: "Online Variance"

$$\mathbb{V}\mathbf{ar} \;=\; \widetilde{\mathbb{V}\mathbf{ar}} - \mathbf{o}(1)$$

Proof: "Online Refinement"

$$\begin{aligned} \mathbb{V}\mathbf{ar} &= \widetilde{\mathbb{V}\mathbf{ar}} - \mathbf{o}(1) \\ \mathcal{R}^{\mathrm{b}} &= \widetilde{\mathcal{R}}^{\mathrm{b}} - \mathbf{o}(1) \end{aligned}$$

Proof: "Online Refinement"

$$
\begin{aligned}
\mathbb{V}\mathrm{ar} &= \widetilde{\mathbb{V}\mathrm{ar}} - \mathrm{o}(1) \\
\mathcal{R}^{\mathrm{b}} &= \tilde{\mathcal{R}}^{\mathrm{b}} - \mathrm{o}(1) \\
&= \frac{1}{T}\sum_{t=1}^{T} \left\| a_t - \bar{a}_{t-1}(b_t) \right\|^2 - \mathrm{o}(1)
\end{aligned}
$$

Proof: "Online Refinement"

$$\mathbb{V}\mathrm{ar} = \widetilde{\mathbb{V}\mathrm{ar}} - \mathrm{o}(1)$$

$$\mathcal{R}^{\mathrm{b}} = \tilde{\mathcal{R}}^{\mathrm{b}} - \mathrm{o}(1)$$

$$= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \|a_t - \bar{a}_{t-1}(b_t)\|^2}_{\mathcal{B}^{\mathrm{c}}} - \mathrm{o}(1)$$

$$= \mathcal{B}^{\mathrm{c}} - \mathrm{o}(1)$$

$c_t = \bar{a}_{t-1}(b_t)$

Calibeating

Calibeating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Self-Calibeating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Theorem

$$c_t = \bar{a}^{\mathrm{c}}_{t-1}(c_t)$$

GUARANTEES c-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{c}} - \mathcal{K}^{\mathrm{c}}$$

Self-Calibeating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leqq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Theorem

$$c_t = \bar{a}^{\mathrm{c}}_{t-1}(c_t)$$

GUARANTEES c-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leqq \mathcal{B}^{\mathrm{c}} - \mathcal{K}^{\mathrm{c}}$$

$$\Leftrightarrow \quad \mathcal{K}^{\mathrm{c}} = 0$$

Self-Calibeating = Calibrating

Theorem

$$c_t = \bar{a}^{\mathrm{b}}_{t-1}(b_t)$$

GUARANTEES b-CALIBEATING:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{b}} - \mathcal{K}^{\mathrm{b}}$$

Theorem

$$c_t = \bar{a}^{\mathrm{c}}_{t-1}(c_t)$$

GUARANTEES CALIBRATION:

$$\mathcal{B}^{\mathrm{c}} \leq \mathcal{B}^{\mathrm{c}} - \mathcal{K}^{\mathrm{c}}$$

$$\Leftrightarrow \quad \mathcal{K}^{\mathrm{c}} = 0$$

“Fixed Point”

How do we get $\boldsymbol{c}_t$ “close to” $\bar{\boldsymbol{a}}_{t-1}(\boldsymbol{c}_t)$?

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $g : D \to \mathbb{R}^m$ **arbitrary** function

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $g : D \longrightarrow \mathbb{R}^m$ **arbitrary** function

Theorem There exists a probability distribution P on the δ-grid D of C such that

“Fixed Point”

How do we get c_t “close to” $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $g : D \rightarrow \mathbb{R}^m$ **arbitrary** function

Theorem There exists a probability distribution P on the δ-grid D of C such that

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - g(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Stochastic "Fixed Point"

How do we get c_t "close to" $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $g : D \longrightarrow \mathbb{R}^m$ **arbitrary** function

Theorem There exists a probability distribution P on the δ-grid D of C such that

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - g(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Stochastic "Fixed Point"

How do we get c_t "close to" $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $g : D \longrightarrow \mathbb{R}^m$ **arbitrary** function

Theorem There exists a probability distribution P on the δ-grid D of C such that

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - g(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Obtained by solving a Minimax problem (LP)

Outgoing Minimax (FH)

How do we get c_t "close to" $\bar{a}_{t-1}(c_t)$?

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite δ-grid of C (for $\delta > 0$)
- $g : D \longrightarrow \mathbb{R}^m$ **arbitrary** function

Theorem There exists a probability distribution P on the δ-grid D of C such that

$$\mathbb{E}_{x \sim P}\left[\|v - x\|^2 - \|v - g(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

Obtained by solving a Minimax problem (LP)

Stochastic “Fixed Point” (FH)

Theorem There exists a probability distribution P on the δ-grid D of C such that

$$\mathbb{E}_{x\sim P}\left[\|v - x\|^2 - \|v - g(x)\|^2\right] \leq \delta^2 \quad \forall v \in C$$

- **Obtained by solving a MINIMAX problem (LP)**

Stochastic "Fixed Point" (FH)

Theorem There exists a probability distribution P on the δ-grid D of C such that

$$\mathbb{E}_{x\sim P}\left[\|v-x\|^2-\|v-g(x)\|^2\right]\le\delta^2\quad\forall v\in C$$

- **Obtained by solving a MINIMAX problem (LP)**
- Moreover: solving a **FIXED POINT** problem yields a probability distribution η that is **ALMOST DETERMINISTIC**: its support is included in a ball of size δ

Calibrating

Calibrating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBRATION**

Calibrating

Theorem

There is a stochastic procedure that **GUARANTEES** **CALIBRATION**

Proof. Self-calibeating + Outgoing Minimax

Calibrating

Theorem

There is a stochastic procedure that **GUARANTEES** **CALIBRATION**

Proof. Self-calibeating + Outgoing Minimax

Note. δ-**CALIBRATION**

Calibrated Calibeating

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES CALIBEATING**

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES** **CALIBEATING** and **CALIBRATION**

Calibrated Calibeating

Theorem

There is a stochastic procedure that **GUARANTEES** **CALIBEATING** and **CALIBRATION**

Proof. Calibeat the **joint** binning of $\boldsymbol{b}$ and $\boldsymbol{c}$, by the Outgoing Minimax theorem

Continuous-Calibrated Calibeating

Continuous-Calibrated Calibeating

Theorem

There is a ***deterministic*** procedure that **GUARANTEES** **CALIBEATING**

Continuous-Calibrated Calibeating

Theorem

There is a ***deterministic*** procedure
that **GUARANTEES** **CALIBEATING**
and **CONTINUOUS CALIBRATION**

Continuous-Calibrated Calibeating

Theorem

There is a ***deterministic*** procedure that **GUARANTEES** **CALIBEATING** and **CONTINUOUS CALIBRATION**

Proof. Calibeat the **joint** binning of $\boldsymbol{b}$ and $\boldsymbol{c}$, by the Outgoing Fixed Point theorem

Multi-Calibeating

Multi-Calibeating

Theorem

There is a ***deterministic*** procedure that **GUARANTEES**

simultaneous CALIBEATING
of several forecasters

Multi-Calibeating

Theorem

There is a ***stochastic*** procedure

that **GUARANTEES**

simultaneous CALIBEATING

of several forecasters

and **CALIBRATION**

Multi-Calibeating

Theorem

There is a ***stochastic*** procedure
that **GUARANTEES**

simultaneous CALIBEATING
of several forecasters

and **CALIBRATION**

Proof. Calibeat the **joint** binning

In all the results above:

In all the results above:

	CALIBRATION	
Obtained by	***Minimax***	
Procedure	***stochastic***	

... and Continuous Calibration

In all the results above:

	CALIBRATION	CONTINUOUS CALIBRATION
Obtained by	***Minimax***	***Fixed Point***
Procedure	***stochastic***	***deterministic***

Successful Economic Forecasting

Successful Economic Forecasting

TAKING PRIDE IN OUR RECORD

Successful Economic Forecasting

TAKING PRIDE IN OUR RECORD

“We have correctly forecasted 8 of the last 5 recessions”

Successful Economic Forecasting

TAKING PRIDE IN OUR RECORD

"We have correctly forecasted 8 of the last 5 recessions"