“Calibeating”: Beating Forecasters at Their Own Game

Sergiu Hart

June 2021
“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
Dean P. Foster and Sergiu Hart
“Forecast-Hedging and Calibration”
*Journal of Political Economy* (forthcoming)

www.ma.huji.ac.il/hart/abs/calib-int.html
Dean P. Foster and Sergiu Hart
“Forecast-Hedging and Calibration”
*Journal of Political Economy* (forthcoming)

[www.ma.huji.ac.il/hart/abs/calib-int.html](http://www.ma.huji.ac.il/hart/abs/calib-int.html)

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

February 2020 (not yet distributed)
Forecaster says: “*The probability of rain tomorrow is* $p$”
Forecaster says: “The probability of rain tomorrow is $p$”

Forecaster is CALIBRATED if
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 can be guaranteed
(no matter what the weather will be)
Calibration can be guaranteed (no matter what the weather will be)

Foster and Vohra 1994 [publ 1998]
**CALIBRATION** can be guaranteed
(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
CALIBRATION can be guaranteed
(no matter what the weather will be)

- Foster and Vohra 1994 [publ 1998]
- Hart 1995: proof by Minimax Theorem
- . . .
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 2021/22]: simplest procedure, by “Forecast-Hedging”
AVERAGE ACTION (= frequency of rain)
AVERAGE ACTION (= frequency of rain)

perfect calibration
Forecast-Hedging

AVERAGE ACTION (= frequency of rain)

perfect calibration
AVERAGE ACTION (= frequency of rain)

perfect calibration
AVERAGE ACTION (= frequency of rain)

perfect calibration
AVERAGE ACTION (= frequency of rain)

perfect calibration
Calibration in Practice

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

Betting vs Reality

Percent that it actually happened

Prediction buckets

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
</table>
## Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
| time | 1  | 2  | 3  | 4  | 5  | 6  | ...
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>...</td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

**F1:** \(\text{CALIBRATION} = 0\)
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

**F1:** \(\text{CALIBRATION} = 0\)

**F2:** \(\text{CALIBRATION} = 0\)
<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

F1: \( \text{CALIBRATION} = 0 \) \( \text{IN-BIN VARIANCE} = 0 \)
F2: \( \text{CALIBRATION} = 0 \)
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  **IN-BIN VARIANCE** = $\frac{1}{4}$
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 \( x \)

\[
\bar{a}(x) = \frac{\sum_{t=1}^{T} 1_{x(c_t)} a_t}{\sum_{t=1}^{T} 1_{x(c_t)}}
\]
Notations

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

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

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

$$\mathcal{K} = \frac{1}{T} \sum_{t=1}^{T} \| c_t - \bar{a}(c_t) \|^2$$
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 \( x \)
- \( \mathcal{K} \equiv \mathcal{K}_T = \text{CALIBRATION score} = \) average distance between \( c_t \) and \( \bar{a}(c_t) \)
Notations

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

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

$$
\mathcal{R} = \frac{1}{T} \sum_{t=1}^{T} \|a_t - \bar{a}(c_t)\|^2
$$
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
$\mathcal{B} \equiv \mathcal{B}_T = \text{BRIER}$ (1950) score = average distance between $a_t$ and $c_t$
Brier score

\[ \mathcal{B} \equiv \mathcal{B}_T = \textbf{BRIER} \ (1950) \text{ score } = \text{ average distance between } a_t \text{ and } c_t \]

\[ \mathcal{B} = \frac{1}{T} \sum_{t=1}^{T} \| a_t - c_t \|^2 \]
Brier score

\[ B \equiv B_T = \text{BRIER} \text{ (1950) score} = \text{average distance between } a_t \text{ and } c_t \]
\( \mathcal{B} \equiv \mathcal{B}_T = \text{BRIER} \) (1950) score = average distance between \( a_t \) and \( c_t \)

\[ \mathcal{B} = \mathcal{R} + \mathcal{K} \]
Brier score

\[ B \equiv B_T = BRIER \] (1950) score = average distance between \( a_t \) and \( c_t \)

\[ B = R + K \]

BRIER = RENFINEMENT + CALIBRATION
Brier score

\[ B \equiv B_T = \text{Brier} \ (1950) \ \text{score} = \text{average distance between} \ a_t \ \text{and} \ c_t \]

\[ B = R + K \]

\text{Brier} = \text{REFINEMENT} + \text{CALIBRATION}

Proof.

\[ \mathbb{E}[(X - c)^2] = \text{Var}(X) + (\bar{X} - c)^2 \]

where \( c \) is a constant and \( X \) is a random variable with \( \bar{X} = \mathbb{E}[X] \)
Brier score

\[ B \equiv B_T = \text{Brier} \text{ (1950) score} = \text{average distance between} \ a_t \ \text{and} \ c_t \]

\[ B = R + K \]

\text{Brier} = \text{Refinement} + \text{Calibration}
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

F1: \( \text{CALIBRATION} = 0 \) \( \text{IN-BIN VARIANCE} = 0 \)

F2: \( \text{CALIBRATION} = 0 \) \( \text{IN-BIN VARIANCE} = \frac{1}{4} \)
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

**F1:** \( K = 0 \quad R = 0 \)

**F2:** \( K = 0 \quad R = \frac{1}{4} \)
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>F1</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td>50%</td>
<td></td>
</tr>
</tbody>
</table>

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”
Testing experts:
Testing experts: 
✓ BRIER score
Testing experts:

- BRIER score
- CALIBRATION score
“Experts”
Experts better be **CALIBRATED**
(since calibration can be **guaranteed**)
“Experts”

- Experts better be **CALIBRATED**
  (since calibration can be *guaranteed*)
- One can always **CALIBRATE** an expert:
Experts better be **CALIBRATED** (since calibration can be **guaranteed**)

One can always **CALIBRATE** an expert:

- Replace each $c_t$ with $\bar{a}_T(c_t)$
Experts better be **CALIBRATED**
(since calibration can be **guaranteed**)

One can always **CALIBRATE** an expert:

- Replace each $c_t$ with $\bar{a}_T(c_t)$

$$\Rightarrow \mathcal{K}' = 0 \quad \mathcal{R}' = \mathcal{R}$$
Experts better be **CALIBRATED**
(since calibration can be **guaranteed**)

One can always **CALIBRATE** an expert:

- Replace each $c_t$ with $\bar{a}_T(c_t)$

  $\Rightarrow K' = 0$  \hspace{1cm} $R' = R$

  $\Rightarrow B' = B - K$
Experts better be \textbf{CALIBRATED} (since calibration can be \textit{guaranteed})

One can always \textbf{CALIBRATE} an expert:

Replace each $c_t$ with $\bar{a}_T(c_t)$

$\Rightarrow K' = 0 \quad R' = R$

$\Rightarrow B' = B - K$

\underline{EX POST / OFFLINE}

(when the $\bar{a}_T(c_t)$ are known)
Experts better be **CALIBRATED** (since calibration can be **guaranteed**)

One can always **CALIBRATE** an expert:

- Replace each \( c_t \) with \( \bar{a}_T(c_t) \)

\[
\begin{align*}
K' &= 0 \\
R' &= R \\
B' &= B - K
\end{align*}
\]

**EX POST / OFFLINE**

(when the \( \bar{a}_T(c_t) \) are known)

**Question:** *Can one do it **ONLINE**?*
Experts better be **CALIBRATED**
(since calibration can be **guaranteed**)

One can always **CALIBRATE** an expert:

- Replace each $c_t$ with $\bar{a}_T(c_t)$

\[ \Rightarrow K' = 0 \quad R' = R \]
\[ \Rightarrow B' = B - K \]

**EX POST / OFFLINE**
(when the $\bar{a}_T(c_t)$ are known)

---

**Question:** *Can one do it **ONLINE**?*

without losing the **REFINEMENT**?
Consider a forecasting sequence $b_t$ (in a finite set $B$)
Consider a forecasting sequence $b_t$ (in a finite set $B$)

At each time $t$ generate a forecast $c_t$
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 $b_t$ (in a finite set $B$)

At each time $t$ generate a forecast $c_t$

**ONLINE**: use only $b_t$ and history

such that

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

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

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

\[
B_c^c \leq B^b - \mathcal{K}^b
\]
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

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

\[ B^c \leq B^b - K^b = R^b \]

$c$ “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

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

$c$ “BEATS” $b$ by $b$’s CALIBRATION score

GUARANTEED for ALL sequences of actions and forecasts
Example
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( b )</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td></td>
</tr>
</tbody>
</table>

\[ b: \quad \mathcal{K}^b = 0.1 \quad \mathcal{R}^b = 0 \quad \mathcal{B}^b = 0.1 \]
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

\[ b: \, K^b = 0.1 \quad R^b = 0 \quad B^b = 0.1 \]
### Example

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

\[ b: \quad \mathcal{K}^b = 0.1 \quad \mathcal{R}^b = 0 \quad \mathcal{B}^b = 0.1 \]

\[ c: \quad \mathcal{K}^c = 0 \quad \mathcal{R}^c = 0 \quad \mathcal{B}^c = 0 \]
### Calibeating

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>b</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>c</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

- **b:** $K^b = 0.1$  
  $R^b = 0$  
  $B^b = 0.1$

- **c:** $K^c = 0$  
  $R^c = 0$  
  $B^c = 0$

### c calibeats b

$c$ calibeats $b$:  
$B^c \leq B^b - K^b$
Calibeating

<table>
<thead>
<tr>
<th>time</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>rain</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(b)</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td>80%</td>
<td>40%</td>
<td></td>
</tr>
<tr>
<td>(c)</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

\(b\): \(\mathcal{K}^b = 0.1\) \(\mathcal{R}^b = 0\) \(\mathcal{B}^b = 0.1\)

\(c\): \(\mathcal{K}^c = 0\) \(\mathcal{R}^c = 0\) \(\mathcal{B}^c = 0\)

\(c\) *calibeating* \(b\): \(\mathcal{B}^c \leq \mathcal{B}^b - \mathcal{K}^b = \mathcal{R}^b\)
Calibeating

(that was easy ...)

Calibating

(that was easy ...)

*Can one calibrate 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
Calibearing

(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 ($\mathcal{R}^b > 0$)
(that was easy ...)

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

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

(that was easy ...)

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

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

(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 ($R^b > 0$)
- **Bin averages** do not converge
- **ONLINE**
- **GUARANTEED** (even against adversary)
A Simple Way to Calibrate
Theorem

The procedure

\[ c_t = \bar{a}^b_{t-1}(b_t) \]

GUARANTEES \text{ b-CALIBEATING}
A Simple Way to Calibeat

Theorem

The procedure

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

GUARANTEES b-CALIBEATING

Forecast the average action of the current b-forecast
Proof

\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \]
Proof

\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \]

\[ = \frac{1}{T} \sum_{t=1}^{T} \left( 1 - \frac{1}{t} \right) \| x_t - \bar{x}_{t-1} \|^2 \]
\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \left\| x_t - \bar{x}_T \right\|^2 \]

\[ = \frac{1}{T} \sum_{t=1}^{T} \left( 1 - \frac{1}{t} \right) \left\| x_t - \bar{x}_{t-1} \right\|^2 \]
\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \\
= \frac{1}{T} \sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \| x_t - \bar{x}_{t-1} \|^2 \]
\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \]

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

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

\[
\text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \\
= \frac{1}{T} \sum_{t=1}^{T} \left(1 - \frac{1}{t}\right) \| x_t - \bar{x}_{t-1} \|^2 \\
= \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_{t-1} \|^2 - o(1)
\]

(*) \hspace{1cm} o(1) = O \left(\frac{\log T}{T}\right)
Proof

\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \]

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

\[ = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_{t-1} \|^2 - o(1) \]
Proof: “Online Variance"

\[ \text{Var} = \frac{1}{T} \sum_{t=1}^{T} \| x_t - \bar{x}_T \|^2 \]

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

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

\[ = \hat{\text{Var}} - o(1) \]
Proof: “Online Variance”

\[ \text{Var} = \widetilde{\text{Var}} - o(1) \]
Proof: “Online Refinement”

\[ \text{Var} = \tilde{\text{Var}} - o(1) \]

\[ \mathcal{R}^b = \tilde{\mathcal{R}}^b - o(1) \]
Proof: “Online Refinement”

\[ \text{Var} = \tilde{\text{Var}} - o(1) \]

\[ \mathcal{R}^b = \tilde{\mathcal{R}}^b - o(1) \]

\[ = \frac{1}{T} \sum_{t=1}^{T} \| a_t - \tilde{a}_{t-1}(b_t) \|^2 - o(1) \]
Proof: “Online Refinement”

\[ \text{Var} = \tilde{\text{Var}} - o(1) \]

\[ \mathcal{R}^b = \tilde{\mathcal{R}}^b - o(1) \]

\[ = \frac{1}{T} \sum_{t=1}^{T} \| a_t - \bar{a}_{t-1}(b_t) \|^2 - o(1) \]

\[ = \mathcal{B}^c - o(1) \]

\[ c_t = \bar{a}_{t-1}(b_t) \]
Calibeating

Theorem

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

GUARANTEES b-CALIBEATING:

\[ \mathcal{B}_c \leq \mathcal{B}_b - \mathcal{K}_b \]
Self-Calibeating

Theorem
\[ c_t = \bar{a}^b_{t-1}(b_t) \]

GUARANTEES \textit{b-CALIBEATING}:
\[ B^c \leq B^b - K^b \]

Theorem
\[ c_t = \bar{a}^c_{t-1}(c_t) \]

GUARANTEES \textit{c-CALIBEATING}:
\[ B^c \leq B^c - K^c \]
Self-Calibeating

**Theorem**

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

**GUARANTEES \(b\)-CALIBEATING:**

\[ B^c \leq B^b - K^b \]

---

**Theorem**

\[ c_t = \bar{a}_{t-1}^c(c_t) \]

**GUARANTEES \(c\)-CALIBEATING:**

\[ B^c \leq B^c - K^c \]

\[ \Leftrightarrow K^c = 0 \]
Self-Calibeating = Calibrating

**Theorem**

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

GUARANTEES \textit{b-CALIBEATING}:

\[ \mathcal{B}^c \leq \mathcal{B}^b - \mathcal{K}^b \]

**Theorem**

\[ c_t = \bar{a}_{t-1}^c(c_t) \]

GUARANTEES \textit{CALIBRATION}:

\[ \mathcal{B}^c \leq \mathcal{B}^c - \mathcal{K}^c \]

\[ \iff \mathcal{K}^c = 0 \]
“Fixed Point”

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

**Theorem** There exists a probability distribution on (a $\delta$-grid $D$ of) $C$ such that for every $x \in C$
Stochastic “Fixed Point”

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

**Theorem** There exists a probability distribution on (a $\delta$-grid $D$ of) $C$ such that for every $x \in C$

$$\mathbb{E}_c \left[ \|x - c\|^2 - \|x - g(c)\|^2 \right] \leq \delta^2$$
How do we get $c_t$ “close to” $\bar{a}_{t-1}(c_t)$?

**Theorem** There exists a probability distribution on (a $\delta$-grid $D$ of) $C$ such that for every $x \in C$

$$\mathbb{E}_c \left[ \|x - c\|^2 - \|x - g(c)\|^2 \right] \leq \delta^2$$

- $C \subset \mathbb{R}^m$ compact convex
- $D \subset C$ finite $\delta$-grid of $C$ for $\delta > 0$
- $g : D \rightarrow \mathbb{R}^m$ arbitrary function
Stochastic “Fixed Point”

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

**Theorem** There exists a probability distribution on (a $\delta$-grid $D$ of) $C$ such that for every $x \in C$

$$\mathbb{E}_c \left[ \| x - c \|^2 - \| x - g(c) \|^2 \right] \leq \delta^2$$

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

**Obtained by solving a Minimax problem (LP)**
How do we get $c_t$ “close to” $\bar{a}_{t-1}(c_t)$?

**Theorem** There exists a probability distribution on (a $\delta$-grid $D$ of) $C$ such that for every $x \in C$

$$
\mathbb{E}_c \left[ \|x - c\|^2 - \|x - g(c)\|^2 \right] \leq \delta^2
$$

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

**Obtained by solving a Minimax problem (LP)**
Theorem There exists a probability distribution on (a \( \delta \)-grid \( D \) of) \( C \) such that for every \( x \in C \)

\[
\mathbb{E}_c \left[ \| x - c \|^2 - \| x - g(c) \|^2 \right] \leq \delta^2
\]

• Obtained by solving a Minimax problem (LP)
Stochastic “Fixed Point” (FH)

**Theorem** There exists a probability distribution on (a $\delta$-grid $D$ of) $C$ such that for every $x \in C$

\[
\mathbb{E}_c \left[ \| x - c \|^2 - \| x - g(c) \|^2 \right] \leq \delta^2
\]

- Obtained by solving a Minimax problem (LP)

- Moreover, solving a Fixed Point problem yields a probability distribution that is **ALMOST DETERMINISTIC**: its support is included in a ball of size $\delta$
There is a stochastic procedure that GUARANTEES CALIBRATION
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.* $\delta$-**CALIBRATION**
Calibrated Calibating
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 $b$ and $c$, by the Outgoing Minimax theorem
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
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:

<table>
<thead>
<tr>
<th></th>
<th>CALIBRATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtained by</td>
<td><em>Minimax</em></td>
</tr>
<tr>
<td>Procedure</td>
<td><em>stochastic</em></td>
</tr>
</tbody>
</table>
... and Continuous Calibration

In all the results above:

<table>
<thead>
<tr>
<th>CALIBRATION</th>
<th>CONTINUOUS CALIBRATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obtained by</td>
<td>Minimax</td>
</tr>
<tr>
<td>Procedure</td>
<td>stochastic</td>
</tr>
</tbody>
</table>
Successful Economic Forecasting
Successful Economic Forecasting

TAKING PRIDE IN OUR RECORD
TAKING PRIDE IN OUR RECORD

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