Point predictions and elicitability

Johanna Ziegel

ETH Zurich

CUSO Summer School
7–10 September 2025

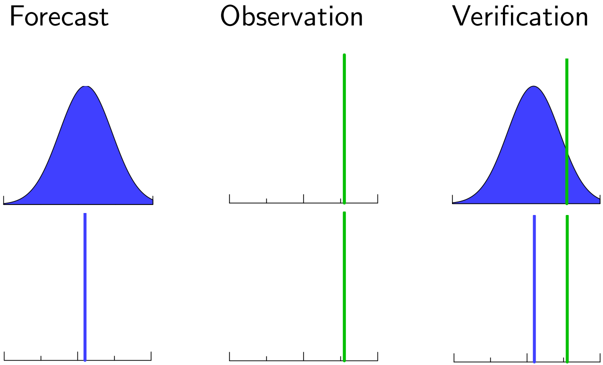
A simple forecasting example

- On Monday, 1 September at 13:50, the weather forecast of MeteoSwiss for Anzère on Sunday, 7 September at 17:00 stated that

  the temperature will be 19.5°C,
  and there is a 3% chance of rain.

- Difference between these two forecasts:
  - Temperature forecast is a point forecast.
  - "Chance of rain" forecast is a probabilistic forecast.
- A probabilistic forecast for temperature could be: $\mathcal{N}(19.5, \sigma^2)$.

Forecasts for real-valued quantities



Some notation

Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a probability space.

- The future event Y is a real valued random variable.
- Let $\mathcal{A} \subseteq \mathcal{F}$ be a sub σ-algebra.
  Our information today: Often specified as covariates Z, so $\mathcal{A} = \sigma(Z)$.

Probabilistic forecast

- Forecast is a probability measure P which is $\mathcal{A}$-measurable.
  (that is, a Markov kernel from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- Optimal prediction: $P = \mathcal{L}(Y|\mathcal{A})$.

Point forecast

- $\mathcal{A}$-measurable random variable $X \in \mathsf{A} \subseteq \mathbb{R}^k$
- Optimal prediction?

Some notation

Let $(\Omega, \mathcal{F}, \mathbb{Q})$ be a probability space.

- The future event Y is a real valued random variable.
- Let $\mathcal{A} \subseteq \mathcal{F}$ be a sub σ-algebra.
  Our information today: Often specified as covariates Z, so $\mathcal{A} = \sigma(Z)$.

Probabilistic forecast

- Forecast is a probability measure P which is $\mathcal{A}$-measurable.
  (that is, a Markov kernel from $(\Omega, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.
- Optimal prediction: $P = \mathcal{L}(Y|\mathcal{A})$.

Point forecast

- $\mathcal{A}$-measurable random variable $X \in \mathsf{A} \subseteq \mathbb{R}^k$
- Optimal prediction?

Outline

Elicitability

- Simulation example
- Definition
- Bregman loss functions
- Quantiles
- Loss functions and scoring rules
- Non-elicitable functionals
- Higher order elicitability

“Best” point forecasts

“Give me your best point prediction x for Y!”

is (usually) not a well-defined task...

Forecast comparison using loss functions

Sequence of realizations $y_1, \dots, y_n \in \mathbb{R}$
Sequences of forecasts $x_{11}, \dots, x_{1n}$, $x_{21}, \dots, x_{2n} \in \mathsf{A}$

Strategy

- Choose a loss function $L : \mathsf{A} \times \mathbb{R} \to \mathbb{R}$
- Compute realized average loss

$$\bar{L}_k = \frac{1}{n} \sum_{j=1}^{n} L_i(x_{kj}, y_j), \quad k = 1, 2$$

- Give preference to forecaster with the lower average loss
- Assess significance of sign of the average loss difference by a Diebold-Mariano test (Diebold and Mariano, 1995; Giacomini and White, 2006; Henzi and Ziegel, 2022; Choe and Ramdas, 2024)

A simulation study

We observe realizations y_t, $t = 1, 2, \ldots$ of a volatile asset

$$Y_t = Z_t^2,$$

where

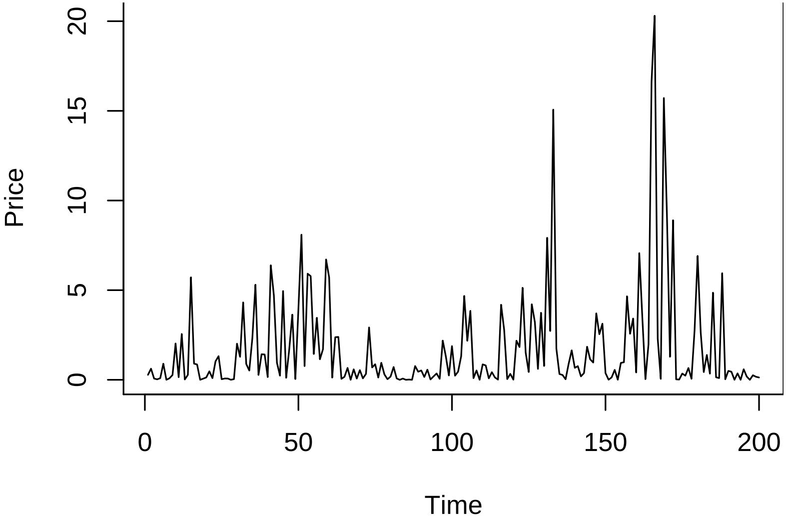
$$Z_t \sim \mathcal{N}(0, \sigma_t^2),$$

and

$$\sigma_t^2 = 0.20\, Z_{t-1}^2 + 0.75\, \sigma_{t-1}^2 + 0.05.$$

(Gneiting, 2011)

A realization



Point forecasts

Statistician

$$\hat{x}_t = \mathbb{E}[Y_t \mid \sigma_t^2] = \sigma_t^2$$

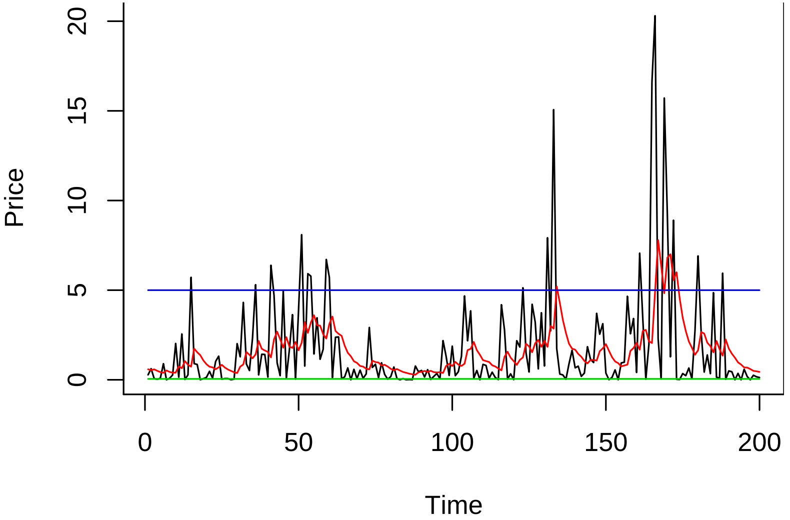
Optimist

$$\hat{x}_t = 5$$

Pessimist

$$\hat{x}_t = 0.05$$

A realization with predictions



Evaluating the point forecasts

Loss functions

$$
\begin{array}{rl}
(x-y)^2 & \text{squared error (SE)} \\
|x-y| & \text{absolute error (AE)} \\
|(x-y)/y| & \text{absolute percentage error (APE)} \\
|(x-y)/x| & \text{relative error (RE)}
\end{array}
$$

Evaluating the point forecasts

Loss functions

$(x-y)^2$	squared error (SE)
$\lvert x-y\rvert$	absolute error (AE)
$\lvert (x-y)/y\rvert$	absolute percentage error (APE)
$\lvert (x-y)/x\rvert$	relative error (RE)

Mean error measures

Forecaster	SE	AE	APE	RE
Statistician	**3.75**	0.95	19796	0.97
Optimist	20.66	4.31	93820	**0.86**
Pessimist	5.31	**0.94**	**939**	18.77

Mr. Bayes

Mr. Bayes issues the optimal point forecasts for the different scoring functions:

$$\hat{x}_t = \arg\min_x \mathbb{E}_{\mathcal{N}(0,\sigma_t^2)} L(x, Z^2)$$

Mr. Bayes

Mr. Bayes issues the optimal point forecasts for the different scoring functions:

$$\hat{x}_t = \arg\min_x \mathbb{E}_{\mathcal{N}(0,\sigma_t^2)} L(x, Z^2)$$

Mean error measures

Forecaster	SE	AE	APE	RE
Statistician	**3.75**	0.95	19796	0.97
Optimist	20.66	4.31	93820	**0.86**
Pessimist	5.31	**0.94**	**939**	18.77
Mr. Bayes	**3.75**	**0.83**	**1.00**	**0.75**

Optimal point forecasts

Loss function is given

Given $L : \mathsf{A} \times \mathcal{Y} \to \overline{\mathbb{R}}$, the optimal prediction is the Bayes rule,

$$X \in \arg\min_{x \in \mathsf{A}} \mathbb{E}(L(x, Y)|\mathcal{A}).$$

Optimal point forecasts

Loss function is given

Given $L : \mathsf{A} \times \mathcal{Y} \to \overline{\mathbb{R}}$, the optimal prediction is the Bayes rule,

$$X \in \arg\min_{x \in \mathsf{A}} \mathbb{E}(L(x, Y)|\mathcal{A}).$$

Functional is given

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$,

$$T : \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P).$$

Examples: mean, median, some risk measure, ...

Optimal point forecasts

Loss function is given

Given $L : \mathsf{A} \times \mathcal{Y} \to \overline{\mathbb{R}}$, the optimal prediction is the Bayes rule,

$$X \in \arg\min_{x \in \mathsf{A}} \mathbb{E}(L(x, Y)|\mathcal{A}).$$

Functional is given

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$,

$$T : \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P).$$

Examples: mean, median, some risk measure, ...
Optimal prediction is

$$X = T(\mathcal{L}(Y|\mathcal{A})).$$

Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$. Let

$$T\colon \mathcal{P} \to \mathrm{A}, \quad P \mapsto T(P)$$

be a functional.

Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$. Let

$$T: \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P)$$

be a functional.

Definition

A loss function $L : \mathsf{A} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is *consistent for T relative to $\mathcal{P}$*, if

$$\mathbb{E}_P L(T(P), Y) \leq \mathbb{E}_P L(x, Y), \quad P \in \mathcal{P}, x \in \mathsf{A}.$$

It is *strictly consistent* if "$=$" implies $x = T(P)$.

Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$. Let

$$T : \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P)$$

be a functional.

Definition

A loss function $L : \mathsf{A} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is *consistent for T relative to $\mathcal{P}$*, if

$$\mathbb{E}_P L(T(P), Y) \leq \mathbb{E}_P L(x, Y), \quad P \in \mathcal{P}, x \in \mathsf{A}.$$

It is *strictly consistent* if "$=$" implies $x = T(P)$.
The functional T is *elicitable relative to $\mathcal{P}$* if there exists a loss function L that is strictly consistent for it.

In other words,

$$T(P) = \arg\min_{x \in \mathsf{A}} \mathbb{E}_P L(x, Y).$$

(Osband, 1985; Lambert et al., 2008; Gneiting, 2011)

Some examples

- Mean: $L(x, y) = (x - y)^2$
  - Least squares regression
  - Comparison of models/forecast performance in terms of MSE
- α-Quantiles (VaR_α): $L(x, y) = (\mathbb{1}\{y \le x\} - \alpha)(x - y)$
  - Quantile regression
- α-Expectiles: $L(x, y) = (\mathbb{1}\{y \le x\} - \alpha)(x - y)^2$
  - Expectile regression (Newey and Powell, 1987)

The mean: Bregman loss functions

Theorem

Let $\mathcal{P}$ be a class of probability measures with finite first moments. Let ϕ be a (strictly) convex function such that $\mathbb{E}_P\phi(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x)$$

is (strictly) consistent for the mean.

The mean: Bregman loss functions

Theorem

Let $\mathcal{P}$ be a class of probability measures with finite first moments. Let ϕ be a (strictly) convex function such that $\mathbb{E}_P\phi(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x)$$

is (strictly) consistent for the mean.

- Under suitable assumptions on $\mathcal{P}$, the *Bregman functions* are the only strictly consistent non-negative loss functions for the mean.

The mean: Bregman loss functions

Theorem

Let $\mathcal{P}$ be a class of probability measures with finite first moments. Let ϕ be a (strictly) convex function such that $\mathbb{E}_P\phi(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x,y) = \phi(y) - \phi(x) - \phi'(x)(y-x)$$

is (strictly) consistent for the mean.

- Under suitable assumptions on $\mathcal{P}$, the *Bregman functions* are the only strictly consistent non-negative loss functions for the mean.
- Choosing $\phi(y) = y^2/(1+|y|)$ shows that the mean is elicitable with respect to the class of all probability measures with finite first moment.

(Savage, 1971)

Proof

$$L(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x)$$

Since ϕ is (strictly) convex, the subgradient inequality states

$$\phi(y) \geq \phi(x) + \phi'(x)(y - x)$$

with equality if (and only if) $x = y$. Therefore,

$$\mathbb{E}_P L(x, Y) - \mathbb{E}_P L(\mathbb{E}_P X, Y) = \phi(\mathbb{E}_P X) - \phi(X) - \phi'(x)(\underbrace{\mathbb{E}_P Y}_{=\mathbb{E}_P X} - x) \geq 0$$

with equality if (and only if) $x = \mathbb{E}_P X$.

Connection to characterization of proper scoring rules

Theorem

Let $\mathcal{P}$ be a class of probability measures with finite first moments. Let ϕ be a (strictly) convex function such that $\mathbb{E}_P\phi(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x)$$

is (strictly) consistent for the mean.

Connection to characterization of proper scoring rules

Theorem

Let $\mathcal{P}$ be a class of probability measures with finite first moments. Let ϕ be a (strictly) convex function such that $\mathbb{E}_P\phi(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x, y) = \phi(y) - \phi(x) - \phi'(x)(y - x)$$

is (strictly) consistent for the mean.

Theorem

A regular scoring rule $S : \mathcal{P} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is (strictly) proper if and only if there is a (strictly) concave function $H : \mathcal{P} \to \mathbb{R}$ such that

$$S(F, y) = H(F) + h_F(y) - \int h_F(x)\, dF(x) = H(F) + \int h_F(x)\, d(\delta_y - F)(x)$$

for every $F \in \mathcal{P}$ and $y \in \mathcal{Y}$, where h_F is a supergradient of H at F.

Quantiles

Theorem

Let $\alpha \in (0,1)$ *and let* $\mathcal{P}$ *be a class of probability measures with unique* α*-quantiles. Let* g *be a (strictly) increasing function such that* $\mathbb{E}_P g(Y)$ *exists and is finite for all* $P \in \mathcal{P}$*. Then,*

$$L(x,y) = (\mathbb{1}\{x \geq y\} - \alpha)(g(x) - g(y))$$

is (strictly) consistent for the α*-quantile.*

Quantiles

Theorem

Let $\alpha \in (0,1)$ and let $\mathcal{P}$ be a class of probability measures with unique α-quantiles. Let g be a (strictly) increasing function such that $\mathbb{E}_P g(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x,y) = (\mathbb{1}\{x \geq y\} - \alpha)(g(x) - g(y))$$

is (strictly) consistent for the α-quantile.

- Under suitable assumptions on $\mathcal{P}$, the *asymmetric piecewise linear functions* are the only strictly consistent loss functions for the α-quantile.

Quantiles

Theorem

Let $\alpha \in (0,1)$ and let $\mathcal{P}$ be a class of probability measures with unique α-quantiles. Let g be a (strictly) increasing function such that $\mathbb{E}_P g(Y)$ exists and is finite for all $P \in \mathcal{P}$. Then,

$$L(x,y) = (\mathbb{1}\{x \geq y\} - \alpha)(g(x) - g(y))$$

is (strictly) consistent for the α-quantile.

- Under suitable assumptions on $\mathcal{P}$, the *asymmetric piecewise linear functions* are the only strictly consistent loss functions for the α-quantile.
- Choosing g bounded and strictly increasing shows that the α-quantile is elicitable with respect to the class of all probability measures with unique α-quantiles.

(Thomson, 1979)

Consistent loss functions and proper scoring rules

Theorem

Let $T : \mathcal{P} \to \mathsf{A}$ be an elicitable functional with consistent loss function L. Then,

$$S : \mathcal{P} \times \mathcal{Y} \to \overline{R}, \quad P \mapsto S(P, y) = L(T(P), y)$$

is a proper scoring rule, which is typically not strictly proper.

Necessary condition for elicitability

Theorem

If a functional T is elicitable, then it has convex level sets: For $P, Q \in \mathcal{P}$, $\alpha \in (0,1)$ with $(1-\alpha)P + \alpha Q \in \mathcal{P}$ it holds that

$$t = T(P) = T(Q) \implies t = T((1-\alpha)P + \alpha Q).$$

Necessary condition for elicitability

Theorem

If a functional T is elicitable, then it has convex level sets: For $P, Q \in \mathcal{P}$, $\alpha \in (0,1)$ with $(1-\alpha)P + \alpha Q \in \mathcal{P}$ it holds that

$$t = T(P) = T(Q) \implies t = T((1-\alpha)P + \alpha Q).$$

Proof.

Let $x \in A$. Then

$$\begin{aligned}\mathbb{E}_{(1-\alpha)P+\alpha Q} L(t, Y) &= (1-\alpha)\mathbb{E}_P L(t, Y) + \alpha \mathbb{E}_Q L(t, Y) \\ &\leq (1-\alpha)\mathbb{E}_P L(x, Y) + \alpha \mathbb{E}_Q L(x, Y) = \mathbb{E}_{(1-\alpha)P+\alpha Q} L(x, Y),\end{aligned}$$

hence $t = T((1-\alpha)P + \alpha Q)$. □

(Osband, 1985)

Mean

Let $T = \mathbb{E}$ and $\mathbb{E}_P X = \mathbb{E}_Q X = t$. Then, $\mathbb{E}_{(1-\alpha)P+\alpha Q} X = t$.

Mean

Let $T = \mathbb{E}$ and $\mathbb{E}_P X = \mathbb{E}_Q X = t$. Then, $\mathbb{E}_{(1-\alpha)P+\alpha Q} X = t$.

Variance

Let $T = \mathrm{var}$ and $X \sim P = \mathcal{N}(0,1)$, $Y \sim Q = \mathcal{N}(1,1)$. Then,

$$\mathbb{E}X^2 = 1, \quad \mathbb{E}Y^2 = \mathrm{var}(X) + (\mathbb{E}Y)^2 = 1 + 1 = 2.$$

Mean

Let $T = \mathbb{E}$ and $\mathbb{E}_P X = \mathbb{E}_Q X = t$. Then, $\mathbb{E}_{(1-\alpha)P+\alpha Q} X = t$.

Variance

Let $T = \mathrm{var}$ and $X \sim P = \mathcal{N}(0,1)$, $Y \sim Q = \mathcal{N}(1,1)$. Then,

$$\mathbb{E}X^2 = 1, \quad \mathbb{E}Y^2 = \mathrm{var}(X) + (\mathbb{E}Y)^2 = 1 + 1 = 2.$$

Therefore, for $\alpha \in (0,1)$,

$$\begin{aligned} T((1-\alpha)P + \alpha Q) &= \int x^2 \, \mathrm{d}((1-\alpha)P + \alpha Q)(x) - \left(\int x \, \mathrm{d}((1-\alpha)P + \alpha Q)(x) \right)^2 \\ &= (1-\alpha) + 2\alpha - ((1-\alpha) + \alpha) \\ &= \alpha \neq 1. \end{aligned}$$

Elicitable and non-elicitable functionals

Elicitable

- Mean, moments
- Median, quantiles
- Expectiles

Not elicitable

- Variance
- Expected shortfall (Gneiting, 2011)
- Range Value at Risk/Interquantile expectation

Expected shortfall and Range Value at Risk

Let $Y \in \mathbb{R}$ have distribution P.

Expected shortfall (ES)

Let $\alpha \in (0,1)$. If Y has finite mean, we define

$$\mathrm{ES}_\alpha(P) = \frac{1}{\alpha}\int_0^\alpha q_u(P)\,\mathrm{d}u \quad = \mathbb{E}\big(Y \mid Y \leq q_\alpha(P)\big).$$

Range Value-at-Risk (RVaR)

For $0 < \alpha < \beta < 1$, we define

$$\mathrm{RVaR}_{\alpha,\beta}(P) = \frac{1}{\alpha_2 - \alpha_1}\int_{\alpha_1}^{\alpha_2} q_u(P)\,\mathrm{d}u \quad = \mathbb{E}\big(Y \mid q_{\alpha_1}(P) \leq Y \leq q_{\alpha_2}(P)\big).$$

Why do we care about ES?

- ES is an important (coherent) risk measure in banking, finance and insurance.
- Let $Y \sim P$ be the profit of some financial asset or portfolio.
- We consider α close to zero ($\alpha = 0.01$, $\alpha = 0.025$).
- $\mathrm{VaR}_\alpha(P) = q_\alpha(P)$: With probability α, a loss does not exceed this value
- $\mathrm{ES}_\alpha(P)$: Average size of a loss that exceeds $\mathrm{VaR}_\alpha(P)$

Why do we care about ES?

- ES is an important (coherent) risk measure in banking, finance and insurance.
- Let $Y \sim P$ be the profit of some financial asset or portfolio.
- We consider α close to zero ($\alpha = 0.01$, $\alpha = 0.025$).
- $\mathrm{VaR}_\alpha(P) = q_\alpha(P)$: With probability α, a loss does not exceed this value
- $\mathrm{ES}_\alpha(P)$: Average size of a loss that exceeds $\mathrm{VaR}_\alpha(P)$

Problem

- ES is not elicitable (Gneiting, 2011).

Why do we care about ES?

- ES is an important (coherent) risk measure in banking, finance and insurance.
- Let $Y \sim P$ be the profit of some financial asset or portfolio.
- We consider α close to zero ($\alpha = 0.01$, $\alpha = 0.025$).
- $\mathrm{VaR}_\alpha(P) = q_\alpha(P)$: With probability α, a loss does not exceed this value
- $\mathrm{ES}_\alpha(P)$: Average size of a loss that exceeds $\mathrm{VaR}_\alpha(P)$

Problem

- ES is not elicitable (Gneiting, 2011).

Not a solution...

- Expectiles are elicitable, coherent risk measures.

(Ziegel, 2016; Bellini and Bignozzi, 2015; Delbaen et al., 2016; Wang and Ziegel, 2015)

Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$. Let

$$T : \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P)$$

be a functional.

Definition

A loss function $L : \mathsf{A} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is *consistent for T relative to $\mathcal{P}$*, if

$$\mathbb{E}_P L(T(P), Y) \leq \mathbb{E}_P L(x, Y), \quad P \in \mathcal{P}, x \in \mathsf{A}.$$

It is *strictly consistent* if "=" implies $x = T(P)$.
The functional T is *elicitable relative to $\mathcal{P}$* if there exists a loss function L that is strictly consistent for it.

In other words,

$$T(P) = \arg\min_{x \in \mathsf{A}} \mathbb{E}_P L(x, Y).$$

Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$. Let

$$T : \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P)$$

be a functional. Elicitable if

$$T(P) = \arg\min_{x \in \mathsf{A}} \mathbb{E}_P L(x, Y).$$

- All examples were with $\mathcal{Y} = \mathsf{A} = \mathbb{R}$.

Elicitability

Let $\mathcal{P}$ be a class of probability measures on $\mathcal{Y}$. Let

$$T : \mathcal{P} \to \mathsf{A}, \quad P \mapsto T(P)$$

be a functional. Elicitable if

$$T(P) = \arg\min_{x \in \mathsf{A}} \mathbb{E}_P L(x, Y).$$

- All examples were with $\mathcal{Y} = \mathsf{A} = \mathbb{R}$.
- Useful to consider $\mathsf{A} = \mathcal{Y}^k$: k-elicitability.

Elicitable functionals

1-Elicitable

- Mean, moments
- Median, quantiles
- Expectiles

2-Elicitable

- (Mean, variance)
- (Second moment, variance)
- $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ (Acerbi and Szekely, 2014; Fissler and Ziegel, 2016)

3-Elicitable

- $(\mathrm{VaR}_\alpha, \mathrm{VaR}_\beta, \mathrm{RVaR}_{\alpha,\beta})$ (Fissler and Ziegel, 2021)

***k*-Elicitable**

- Some spectral risk measures together with several quantiles at certain levels (Fissler and Ziegel, 2016)

$T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$

Theorem

Let $\alpha \in (0,1)$, and $\mathrm{A}_0 := \{x \in \mathbb{R}^2 : x_1 \geq x_2\}$. Let $\mathcal{P}$ be a class of probability measures on $\mathbb{R}$ with finite first moments and unique α-quantiles. Any loss function $L\colon \mathrm{A}_0 \times \mathbb{R} \to \mathbb{R}$ of the form

$$\begin{aligned} L(x_1,x_2,y) &= \big(\mathbb{1}\{y \leq x_1\} - \alpha\big)g(x_1) - \mathbb{1}\{y \leq x_1\}g(y) \\ &\quad + \phi'(x_2)\Big(x_2 + \frac{1}{\alpha}(\mathbb{1}\{y \leq x_1\} - \alpha)\frac{x_1}{\alpha} - \mathbb{1}\{y \leq x_1\}\frac{y}{\alpha}\Big) - \phi(x_2) \end{aligned}$$

is consistent for $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ if $\mathbb{1}_{(-\infty,x_1]}g$ is $\mathcal{P}$-integrable and

- *g is increasing and ϕ is increasing and convex.*

It is strictly consistent if, additionally,

- *ϕ is strictly increasing and strictly convex.*

(Fissler and Ziegel, 2016)

$T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$

Theorem (Part 2)

If $T(\mathcal{P}) = \mathrm{A}_0$, the class $\mathcal{P}$ is rich enough and L fulfils some smoothness conditions, all strictly consistent loss functions for T are of the above form (up to equivalence).

$T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$

Theorem (Part 2)

If $T(\mathcal{P}) = \mathsf{A}_0$, the class $\mathcal{P}$ is rich enough and L fulfils some smoothness conditions, all strictly consistent loss functions for T are of the above form (up to equivalence).

Corollary

If the elements of $\mathcal{P}$ have finite first moment and unique α-quantiles, then the pair $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha) \colon \mathcal{P} \to \mathsf{A}_0$ is elicitable.

(Fissler and Ziegel, 2016)

Proof of sufficiency

For **fixed** $x_1 \in \mathbb{R}$, the function

$$\begin{aligned} x_2 \mapsto L(x_1, x_2, y) &= \left(\mathbb{1}\{y \le x_1\} - \alpha\right) g(x_1) - \mathbb{1}\{y \le x_1\} g(y) \\ &+ \phi'(x_2)\Big(x_2 + \underbrace{\left(\mathbb{1}\{y \le x_1\} - \alpha\right)\frac{x_1}{\alpha} - \mathbb{1}\{y \le x_1\}\frac{y}{\alpha}}_{=:L_\alpha(x_1, y)}\Big) - \phi(x_2), \end{aligned}$$

is a Bregman function. $\rightsquigarrow \arg\min_{x_2} \mathbb{E}_P\left[L(x_1, x_2, Y)\right] = -\mathbb{E}_P\left[L_\alpha(x_1, Y)\right]$.

Proof of sufficiency

For **fixed** $x_1 \in \mathbb{R}$, the function

$$\begin{aligned} x_2 \mapsto L(x_1, x_2, y) &= \big(\mathbb{1}\{y \le x_1\} - \alpha\big) g(x_1) - \mathbb{1}\{y \le x_1\} g(y) \\ &+ \phi'(x_2)\Big(x_2 + \underbrace{\big(\mathbb{1}\{y \le x_1\} - \alpha\big)\frac{x_1}{\alpha} - \mathbb{1}\{y \le x_1\}\frac{y}{\alpha}}_{=: L_\alpha(x_1, y)}\Big) - \phi(x_2), \end{aligned}$$

is a Bregman function. $\rightsquigarrow$ $\arg\min_{x_2} \mathbb{E}_P\left[L(x_1, x_2, Y)\right] = -\mathbb{E}_P\left[L_\alpha(x_1, Y)\right]$.

For **fixed** $x_2 \in \mathbb{R}$, the function

$$\begin{aligned} x_1 \mapsto L(x_1, x_2, y) &= \big(\mathbb{1}\{y \le x_1\} - \alpha\big) G_{x_2}(x_1) - \mathbb{1}\{y \le x_1\} G_{x_2}(y) \\ &+ \phi'(x_2) x_2 - \phi(x_2), \\ G_{x_2}(x_1) &= g(x_1) + \phi'(x_2)\frac{x_1}{\alpha} \end{aligned}$$

is a strictly consistent loss function for VaR_α.
$\rightsquigarrow$ $\arg\min_{x_1} \mathbb{E}_P\left[L(x_1, x_2, Y)\right] = \mathrm{VaR}_\alpha(P)$.

Proof of necessity

Osband's principle

- Osband's principle originates from (Osband, 1985) and gives necessary conditions for strictly consistent loss functions.
- It gives a connection between partial derivatives of the expected loss function and an expected identification function.

Proof of necessity

Osband's principle

- Osband's principle originates from (Osband, 1985) and gives necessary conditions for strictly consistent loss functions.
- It gives a connection between partial derivatives of the expected loss function and an expected identification function.

Definition

An *$\mathcal{P}$-identification function* for a functional T is a function $V\colon \mathrm{A} \times \mathcal{Y} \to \mathbb{R}^k$ such that

$$\mathbb{E}_P V(x, Y) = 0 \quad \Longleftrightarrow \quad x = T(P)$$

for all $P \in \mathcal{P}$ and for all $x \in \mathrm{A}$.

Examples:

- Mean: $V(x, y) = x - y$
- α-quantile: $V(x, y) = \mathbb{1}\{y \le x\} - \alpha$.

Osband's Principle

Theorem (Osband's Principle; Fissler and Z (2016))

Let $\mathcal{P}$ be a convex class of probability measures. Let $T : \mathcal{P} \to \mathsf{A} \subseteq \mathbb{R}^k$ be a surjective, elicitable and identifiable functional with $\mathcal{P}$-identification function $V : \mathsf{A} \times \mathcal{Y} \to \mathbb{R}^k$ and a strictly $\mathcal{P}$-consistent loss function $L : \mathsf{A} \times \mathcal{Y} \to \mathbb{R}$. Under some assumptions, there exists a matrix-valued function $h : \operatorname{int}(\mathsf{A}) \to \mathbb{R}^{k \times k}$ such that

$$\nabla_x \mathbb{E}_P L(x, Y) = h(x) \mathbb{E}_P V(x, Y)$$

for all $x \in \operatorname{int}(\mathsf{A})$ and $P \in \mathcal{P}$.

Osband's Principle

Theorem (Osband's Principle; Fissler and Z (2016))

Let $\mathcal{P}$ be a convex class of probability measures. Let $T\colon \mathcal{P} \to \mathsf{A} \subseteq \mathbb{R}^k$ be a surjective, elicitable and identifiable functional with $\mathcal{P}$-identification function $V\colon \mathsf{A} \times \mathcal{Y} \to \mathbb{R}^k$ and a strictly $\mathcal{P}$-consistent loss function $L\colon \mathsf{A} \times \mathcal{Y} \to \mathbb{R}$. Under some assumptions, there exists a matrix-valued function $h\colon \operatorname{int}(\mathsf{A}) \to \mathbb{R}^{k\times k}$ such that

$$\nabla_x \mathbb{E}_P L(x, Y) = h(x)\, \mathbb{E}_P V(x, Y)$$

for all $x \in \operatorname{int}(\mathsf{A})$ and $P \in \mathcal{P}$.

Key idea: Exploit the first order condition of the minimization problem:

$$\rightsquigarrow \nabla_x \mathbb{E}_P L(x, Y) = 0 \qquad \text{for } x = T(P) \text{ for all } P \in \mathcal{P}.$$

"The gradient ∇L is an identification function."

Osband's Principle

(Under some smoothness conditions)
Second order conditions for the minimization problem: The Hessian

$$\nabla_x^2 \left[\mathbb{E}_P L(x, Y)\right] \in \mathbb{R}^{k \times k}$$

must be symmetric for all $x \in \mathrm{A}, P \in \mathcal{P}$, and positive semi-definite at $x = T(P)$.

Osband's Principle

(Under some smoothness conditions)
Second order conditions for the minimization problem: The Hessian

$$\nabla_x^2 \left[\mathbb{E}_P L(x, Y) \right] \in \mathbb{R}^{k \times k}$$

must be symmetric for all $x \in \mathsf{A}, P \in \mathcal{P}$, and positive semi-definite at $x = T(P)$.

- For $k = 1$ the necessary conditions of Osband's principle directly lead to sufficient conditions: For an oriented identification function, choose some $h > 0$ and integrate.

Osband's Principle

(Under some smoothness conditions)
Second order conditions for the minimization problem: The Hessian

$$\nabla_x^2 \left[\mathbb{E}_P L(x, Y)\right] \in \mathbb{R}^{k \times k}$$

must be symmetric for all $x \in \mathsf{A}, P \in \mathcal{P}$, and positive semi-definite at $x = T(P)$.

- For $k = 1$ the necessary conditions of Osband's principle directly lead to sufficient conditions: For an oriented identification function, choose some $h > 0$ and integrate.
- Harder for $k > 1$:
  - Symmetry/positive semi-definiteness of the Hessian imposes (complicated) restrictions on the function h.
  - Even if $x \mapsto \mathbb{E}_P L(x, Y)$ has only one critical point and the Hessian is positive definite there, we can only guarantee a local minimum!
- $\rightsquigarrow$ Generally, we must verify sufficient conditions on a case by case basis.

Application examples of Osband's principle

In Fissler and Ziegel (2016), we considered:

- Functionals with elicitable components (vectors of quantiles, expectiles, ratios of expectations, . . .)
- Spectral risk measures with finitely supported spectral measure
- In particular: $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$

Calibration

Optimal predictions

Probabilistic forecast

- Ideal: $P = \mathcal{L}(Y \mid \mathcal{A}) = \mathcal{L}(Y \mid Z)$
- Auto-calibration: $P = \mathcal{L}(Y \mid P)$ (Take P as information proxy.)

Point forecast

- Ideal: $X = T(Y \mid \mathcal{A}) = T(Y \mid Z)$
- T-calibration: $X = T(Y \mid X)$ (Take X as information proxy.)

Proposition

Auto-calibration implies T-calibration if T is identifiable.

(Gneiting and Resin, 2023)

Predictions for the mean/expectation

Expectation-calibration means that

$$\mathbb{E}[Y \mid X] = X.$$

Predictions for the mean/expectation

Expectation-calibration means that

$$\mathbb{E}[Y \mid X] = X.$$

- If $\mathbb{E}[Y \mid \mathbf{Z} = \mathbf{z}] = g(\mathbf{z})$ and $X = g(\mathbf{Z})$, then X is expectation-calibrated.
- The less information X has the "easier" it is to be calibrated.

Predictions for the mean/expectation

Expectation-calibration means that

$$\mathbb{E}[Y \mid X] = X.$$

- If $\mathbb{E}[Y \mid \mathbf{Z} = \mathbf{z}] = g(\mathbf{z})$ and $X = g(\mathbf{Z})$, then X is expectation-calibrated.
- The less information X has the "easier" it is to be calibrated.
- Reduces to classical notion of calibration for binary outcomes $Y \in \{0, 1\}$ and probability predictions $p \in [0, 1]$.

Assessing expectation-calibration: CORP reliability diagrams

Data: $(x_1, Y_1), \dots, (x_n, Y_n)$ with $x_i, Y_i \in \mathbb{R}$.

Assessing expectation-calibration: CORP reliability diagrams

Data: $(x_1, Y_1), \dots, (x_n, Y_n)$ with $x_i, Y_i \in \mathbb{R}$.
Postulate that

$$x \mapsto \mathbb{E}[Y \mid X = x]$$

is increasing.

Assessing expectation-calibration: CORP reliability diagrams

Data: $(x_1, Y_1),\dots,(x_n, Y_n)$ with $x_i, Y_i \in \mathbb{R}$.
Postulate that

$$x \mapsto \mathbb{E}[Y \mid X = x]$$

is increasing. Assume that $x_1 \le \cdots \le x_n$. Compute isotonic regression $\hat{x}_1^{(iso)}, \dots, \hat{x}_n^{(iso)}$. Then,

$$\hat{x}_i^{(iso)} \approx \mathbb{E}[Y \mid X = x_i] \overset{\text{under calibration}}{=} x_i$$

- Plot $(x_i, \hat{x}_i^{(iso)})$, $i = 1, \dots, n$ and join points by a line.
- Add diagonal.
- Sometimes: Add histogram of $x_1, \dots, x_n$.

(Gneiting and Resin, 2023)

Example

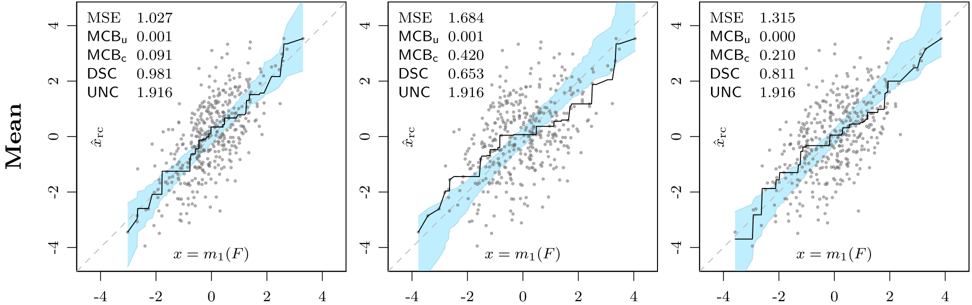


Figure from Gneiting and Resin (2023)

Interval forecasts

- Calibration
- Interval score
- Score decomposition

Interval forecasts

- Uncertainty quantification for an outcome Y is often given in form of a prediction interval $[L, U]$. Conformal prediction...

Interval forecasts

- Uncertainty quantification for an outcome Y is often given in form of a prediction interval $[L, U]$. Conformal prediction...
- Typically it is assessed if
  - marginal coverage is correct/conservative

$$\mathbb{P}(Y \in [L, U]) \geq 1 - \alpha,$$

  - prediction intervals are short on average, that is $\mathbb{E}[|U - L|]$ is small.
- Desirable criteria but unconditional!

Calibration

- Ideally, L is $\alpha/2$-quantile of conditional distribution of Y; U is $(1-\alpha/2)$-quantile of conditional distribution of Y.

Definition

An interval forecast $[L, U]$ is *auto-calibrated* if

$$L = q_{\frac{\alpha}{2}}(Y \mid L, U), \quad U = q_{1-\frac{\alpha}{2}}(Y \mid L, U)$$

$$\mathbb{P}(Y < L \mid L, U) = \frac{\alpha}{2}, \quad \mathbb{P}(Y > L \mid L, U) = \frac{\alpha}{2}.$$

- Auto-calibration implies correct coverage but is more stringent requirement.

Assume unique quantiles for simplicity.

(Allen et al., 2025)

Winkler score

- Interval forecasts are typically assessed using the interval score, or Winkler score:

$$
\begin{aligned}
\mathrm{IS}_\alpha([\ell, u], y) &= |u - \ell| + \frac{2}{\alpha}\mathbb{1}\{y < \ell\}(\ell - y) + \frac{2}{\alpha}\mathbb{1}\{y > u\}(y - u) \\
&= \frac{\alpha}{2}\left[\mathrm{QS}_{\alpha/2}(\ell, y) + \mathrm{QS}_{1-\alpha/2}(u, y)\right].
\end{aligned}
$$

- Winkler score is consistent for the central $1 - \alpha$ prediction interval
- General (non-central) prediction intervals are not elicitable. (Fissler et al., 2021; Brehmer and Gneiting, 2021)

Assessing conditional calibration and discrimination ability

Decompose the expected interval score into miscalibration (MCB), discrimination ability (DSC), and uncertainty (UNC):

$$\mathbb{E}[\mathrm{IS}_\alpha([L,U],Y)] = \underbrace{\mathbb{E}[\mathrm{IS}_\alpha([L,U],Y)] - \mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y \mid L,U), q_{1-\frac{\alpha}{2}}(Y \mid L,U)],Y)]}_{=\mathrm{MCB}}$$
$$- \big(\underbrace{\mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y), q_{1-\frac{\alpha}{2}}(Y)],Y)] - \mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y \mid L,U), q_{1-\frac{\alpha}{2}}(Y \mid L,U)],Y)]}_{=\mathrm{DSC}}\big)$$
$$+ \underbrace{\mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y), q_{1-\frac{\alpha}{2}}(Y)],Y)]}_{=\mathrm{UNC}}$$

- MCB: Expected score of prediction minus expected score of best calibrated prediction.

Assessing conditional calibration and discrimination ability

Decompose the expected interval score into miscalibration (MCB), discrimination ability (DSC), and uncertainty (UNC):

$$\mathbb{E}[\mathrm{IS}_\alpha([L,U],Y)] = \underbrace{\mathbb{E}[\mathrm{IS}_\alpha([L,U],Y)] - \mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y \mid L,U), q_{1-\frac{\alpha}{2}}(Y \mid L,U)],Y)]}_{=\mathrm{MCB}}$$
$$- \Big(\underbrace{\mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y), q_{1-\frac{\alpha}{2}}(Y)],Y)] - \mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y \mid L,U), q_{1-\frac{\alpha}{2}}(Y \mid L,U)],Y)]}_{=\mathrm{DSC}}\Big)$$
$$+ \underbrace{\mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y), q_{1-\frac{\alpha}{2}}(Y)],Y)]}_{=\mathrm{UNC}}$$

- MCB: Expected score of prediction minus expected score of best calibrated prediction.
- DSC: Expected score of best marginal prediction minus best calibrated prediction.

Assessing conditional calibration and discrimination ability

Decompose the expected interval score into miscalibration (MCB), discrimination ability (DSC), and uncertainty (UNC):

$$\mathbb{E}[\mathrm{IS}_\alpha([L,U],Y)] = \underbrace{\mathbb{E}[\mathrm{IS}_\alpha([L,U],Y)] - \mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y \mid L,U), q_{1-\frac{\alpha}{2}}(Y \mid L,U)],Y)]}_{=\mathrm{MCB}}$$
$$- \underbrace{\left(\mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y), q_{1-\frac{\alpha}{2}}(Y)],Y)] - \mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y \mid L,U), q_{1-\frac{\alpha}{2}}(Y \mid L,U)],Y)]\right)}_{=\mathrm{DSC}}$$
$$+ \underbrace{\mathbb{E}[\mathrm{IS}_\alpha([q_{\frac{\alpha}{2}}(Y), q_{1-\frac{\alpha}{2}}(Y)],Y)]}_{=\mathrm{UNC}}$$

- MCB: Expected score of prediction minus expected score of best calibrated prediction.
- DSC: Expected score of best marginal prediction minus best calibrated prediction.
- Practical challenge: Estimate the terms MCB, DSC, UNC
- Suggestion: Use isotonic regression.

(Allen et al., 2025)

Data examples

Student-Teacher Achievement Ratio (STAR) in Tennessee, $n = 433$:

	Original			Recalibrated	
STAR	Interval score	Coverage	Length	Coverage	Length
Ridge	0.22	0.87	0.17	0.87 - 0.91	0.17
Local Ridge	0.23	0.89	0.19	0.75 - 0.94	0.15
Neural Net	0.25	0.91	0.20	0.88 - 0.91	0.19
Local Neural Net	0.28	0.89	0.22	0.86 - 0.92	0.18
CQR Neural Net	0.26	0.88	0.19	0.87 - 0.92	0.18
Quantile Neural Net	0.23	0.91	0.20	0.84 - 0.93	0.17

Data examples

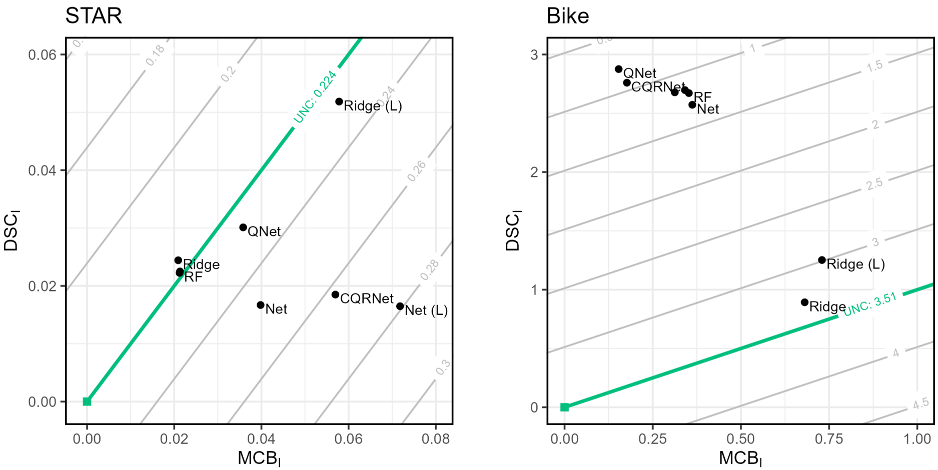
Student-Teacher Achievement Ratio (STAR) in Tennessee, $n = 433$:

	Original			Recalibrated	
STAR	Interval score	Coverage	Length	Coverage	Length
Ridge	0.22	0.87	0.17	0.87 - 0.91	0.17
Local Ridge	0.23	0.89	0.19	0.75 - 0.94	0.15
Neural Net	0.25	0.91	0.20	0.88 - 0.91	0.19
Local Neural Net	0.28	0.89	0.22	0.86 - 0.92	0.18
CQR Neural Net	0.26	0.88	0.19	0.87 - 0.92	0.18
Quantile Neural Net	0.23	0.91	0.20	0.84 - 0.93	0.17

Bike sharing system in Washington D.C., $n = 2178$:

	Original			Recalibrated	
Bike	Interval score	Coverage	Length	Coverage	Length
Ridge	3.30	0.89	2.21	0.88 - 0.92	2.11
Local Ridge	2.99	0.89	2.18	0.81 - 0.93	1.87
Neural Net	1.30	0.89	0.73	0.86 - 0.93	0.67
Local Neural Net	1.15	0.88	0.65	0.85 - 0.93	0.60
CQR Neural Net	0.93	0.90	0.59	0.82 - 0.93	0.58
Quantile Neural Net	0.79	0.90	0.59	0.80 - 0.94	0.49

Data examples



(Allen et al., 2025)

Summary

Summary

- Predictions should be calibrated.
- Predictions should be compared with proper scoring rules or consistent loss functions.
- Scoring rule decompositions allow to understand predictive performance in terms of calibration and discrimination ability.

Thank you for listening and asking questions!

References I

C. Acerbi and B. Szekely. Backtesting expected shortfall. *Risk Magazine*, 27:76–81, 2014.

S. Allen, J. Burnello, and J. Ziegel. Assessing the conditional calibration of interval forecasts using decompositions of the interval score. *Preprint, arXiv: 2508. 18034*, 2025.

F. Bellini and V. Bignozzi. Elicitable risk measures. *Quant. Finance*, 15:725–733, 2015.

J. R. Brehmer and T. Gneiting. Scoring interval forecasts: Equal-tailed, shortest, and modal interval. *Bernoulli*, 27:1993–2010, 2021. doi: 10.3150/20-BEJ1298. URL https://doi.org/10.3150/20-BEJ1298.

Y. J. Choe and A. Ramdas. Comparing sequential forecasters. *Operations Research*, 72: 1368–1387, 2024.

F. Delbaen, F. Bellini, V. Bignozzi, and J. F. Ziegel. Risk measures with the CxLS property. *Finance and Stochastics*, 20:433–453, 2016.

F. X. Diebold and R. S. Mariano. Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13:253–263, 1995.

T. Fissler and J. F. Ziegel. Higher order elicitability and Osband's principle. *Annals of Statistics*, 44:1680–1707, 2016.

References II

T. Fissler and J. F. Ziegel. On the elicitability of range value at risk. *Statistics & Risk Modeling*, 38:25–46, 2021.

T. Fissler, R. Frongillo, J. Hlavinová, and B. Rudloff. Forecast evaluation of quantiles, prediction intervals, and other set-valued functionals. *Electronic Journal of Statistics*, 15: 1034–1084, 2021.

R. Giacomini and H. White. Tests of conditional predictive ability. *Econometrica*, 74: 1545–1578, 2006.

T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106:746–762, 2011.

T. Gneiting and J. Resin. Regression diagnostics meets forecast evaluation: conditional calibration, reliability diagrams, and coefficient of determination. *Electron. J. Stat.*, 17(2): 3226–3286, 2023.

A. Henzi and J. F. Ziegel. Valid sequential inference on probability forecast performance. *Biometrika*, 109:647–663, 2022.

N. Lambert, D. M. Pennock, and Y. Shoham. Eliciting properties of probability distributions. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 129–138, Chicago, Il, USA, 2008. Extended abstract.

References III

W. K. Newey and J. L. Powell. Asymmetric least squares estimation and testing. *Econometrica*, 55:819–847, 1987.

K. H. Osband. *Providing incentives for better cost forecasting*. PhD thesis, University of California, Berkeley, 1985.

L. J. Savage. Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66:783–801, 1971.

W. Thomson. Eliciting production possibilities from a well-informed manager. *Journal of Economic Theory*, 20(3):360–380, 1979.

R. Wang and J. F. Ziegel. Elicitable distortion risk measures: a concise proof. *Statistics and Probability Letters*, 100:172–175, 2015.

J. F. Ziegel. Coherence and elicitability. *Mathematical Finance*, 26:901–918, 2016.