The Elicitation Problem

The Quest of Meaningful Forecast Comparison

Dr. Tobias Fissler

Chapman Fellow at Imperial College London

9 November 2018

Evaluating and Comparing Forecasts

- Forecasts for uncertain future events are ubiquitous and at the heart of strategic decision-making in different contexts:
  - business
  - government
  - risk-management
  - meteorology

Evaluating and Comparing Forecasts

- Forecasts for uncertain future events are ubiquitous and at the heart of strategic decision-making in different contexts:
  - business
  - government
  - risk-management
  - meteorology
- Having m different sources of forecasts, one has the prediction-observation-sequences

$$\big(X_t^{(i)}, Y_t\big)_{t=1,\dots,N} \qquad i = 1, \dots, m.$$

  - $X_t^{(i)} \in \mathsf{A}$ (Action domain). For point forecasts, $\mathsf{A} = \mathbb{R}$ or $\mathsf{A} = \mathbb{R}^k$. For probabilisitic forecasts, $\mathsf{A} = \mathcal{F}$ a space of probability distributions.
  - $Y_t \in \mathsf{O}$ (Observation domain). Usually $\mathsf{O} = \mathbb{R}^d$.

Evaluating and Comparing Forecasts

- Two main tasks:
  - forecast validation
  - forecast comparison/selection.

Evaluating and Comparing Forecasts

- Two main tasks:
  - forecast validation
  - forecast comparison/selection.
- Commonly, one compares competing forecasts in terms of scoring functions (loss functions)

$$S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}, \qquad (x, y) \mapsto S(x, y).$$

Examples: $S(x, y) = |x - y|$, $S(x, y) = (x - y)^2$.

Evaluating and Comparing Forecasts

- Two main tasks:
  - forecast validation
  - forecast comparison/selection.
- Commonly, one compares competing forecasts in terms of scoring functions (loss functions)

$$S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}, \qquad (x, y) \mapsto S(x, y).$$

Examples: $S(x, y) = |x - y|$, $S(x, y) = (x - y)^2$.

- Ranking of the forecasters in terms of realised scores:

$$\mathbf{S}_N^{(1)} = \frac{1}{N}\sum_{t=1}^{N} S\big(X_t^{(1)}, Y_t\big) \overset{?}{\lessgtr} \mathbf{S}_N^{(2)} = \frac{1}{N}\sum_{t=1}^{N} S\big(X_t^{(2)}, Y_t\big)$$

Evaluating and Comparing Forecasts

- Two main tasks:
  - forecast validation
  - forecast comparison/selection.
- Commonly, one compares competing forecasts in terms of scoring functions (loss functions)

$$S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}, \qquad (x, y) \mapsto S(x, y).$$

Examples: $S(x, y) = |x - y|$, $S(x, y) = (x - y)^2$.

- Ranking of the forecasters in terms of realised scores:

$$\mathbf{S}_N^{(1)} = \frac{1}{N}\sum_{t=1}^{N} S\big(X_t^{(1)}, Y_t\big) \overset{?}{\lessgtr} \mathbf{S}_N^{(2)} = \frac{1}{N}\sum_{t=1}^{N} S\big(X_t^{(2)}, Y_t\big)$$

- Ranking depends on the choice of the scoring function!

Evaluating and Comparing Forecasts

- Two main tasks:
  - forecast validation
  - forecast comparison/selection.
- Commonly, one compares competing forecasts in terms of scoring functions (loss functions)

$$S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}, \qquad (x, y) \mapsto S(x, y).$$

Examples: $S(x, y) = |x - y|$, $S(x, y) = (x - y)^2$.

- Ranking of the forecasters in terms of realised scores:

$$\mathbf{S}_N^{(1)} = \frac{1}{N}\sum_{t=1}^{N} S(X_t^{(1)}, Y_t) \overset{?}{\lesseqgtr} \mathbf{S}_N^{(2)} = \frac{1}{N}\sum_{t=1}^{N} S(X_t^{(2)}, Y_t)$$

- Ranking depends on the choice of the scoring function!
- One should disclose the specific choice of the scoring function to the forecasters *ex ante*.

Evaluating and Comparing Forecasts

- Two main tasks:
  - forecast validation
  - forecast comparison/selection.
- Commonly, one compares competing forecasts in terms of scoring functions (loss functions)

$$S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}, \qquad (x, y) \mapsto S(x, y).$$

Examples: $S(x,y) = |x-y|$, $S(x,y) = (x-y)^2$.

- Ranking of the forecasters in terms of realised scores:

$$\mathbf{S}_N^{(1)} = \frac{1}{N}\sum_{t=1}^{N} S(X_t^{(1)}, Y_t) \overset{?}{\lessgtr} \mathbf{S}_N^{(2)} = \frac{1}{N}\sum_{t=1}^{N} S(X_t^{(2)}, Y_t)$$

- Ranking depends on the choice of the scoring function!
- One should disclose the specific choice of the scoring function to the forecasters *ex ante*.
- $\rightsquigarrow$ We need guidance in the choice of the scoring function.

Consistency and Elicitability

- Specification in terms of
  (i) an intrinsically meaningful scoring function (reflecting the actual economic costs); or
  (ii) a property (mean, median, variance, a risk measure) of the underlying distributions of the observation Y_t.
  That is a functional $T\colon \mathcal{F} \to \mathsf{A}$. Here, $\mathcal{F}$ is a class of potential distributions of the observations.

Consistency and Elicitability

- Specification in terms of
  (i) an intrinsically meaningful scoring function (reflecting the actual economic costs); or
  (ii) a property (mean, median, variance, a risk measure) of the underlying distributions of the observation Y_t.
  That is a functional $T \colon \mathcal{F} \to \mathsf{A}$. Here, $\mathcal{F}$ is a class of potential distributions of the observations.
- Using a **Law of Large Numbers** (under ergodicity / mixing assumptions), a forecaster wants to minimise their expected score

$$\mathbf{E}_F[S(x, Y)], \qquad \text{where } Y \sim F.$$

Consistency and Elicitability

- Specification in terms of
  (i) an intrinsically meaningful scoring function (reflecting the actual economic costs); or
  (ii) a property (mean, median, variance, a risk measure) of the underlying distributions of the observation Y_t.
  That is a functional $T\colon \mathcal{F} \to \mathsf{A}$. Here, $\mathcal{F}$ is a class of potential distributions of the observations.
- Using a **Law of Large Numbers** (under ergodicity / mixing assumptions), a forecaster wants to minimise their expected score

$$\mathbf{E}_F\,[S(x, Y)], \qquad \text{where } Y \sim F.$$

- The scoring function should be "unbiased", incentivising truthful forecasts.

Elicitability

Definition 1 (Consistency)

A scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ is strictly $\mathcal{F}$-consistent for some functional $T\colon \mathcal{F} \to \mathsf{A}$ if

$$\mathbf{E}_F[S(T(F), Y)] < \mathbf{E}_F[S(x, Y)]$$

for any $F \in \mathcal{F}$ and any $x \in \mathsf{A}$, $x \neq T(F)$.

Elicitability

Definition 1 (Consistency)

A scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ is strictly $\mathcal{F}$-consistent for some functional $T\colon \mathcal{F} \to \mathsf{A}$ if

$$\mathbf{E}_F[S(T(F), Y)] < \mathbf{E}_F[S(x, Y)]$$

for any $F \in \mathcal{F}$ and any $x \in \mathsf{A}$, $x \neq T(F)$.

Definition 2 (Elicitability)

A functional $T\colon \mathcal{F} \to \mathsf{A}$ is elicitable if there is a strictly $\mathcal{F}$-consistent scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ for T. Then

$$T(F) = \arg\min_{x \in \mathsf{A}} \mathbf{E}_F[S(x, Y)].$$

Relevance and Applications

- **Mathematical / Statistical Applications:**
  - Forecast evaluation, comparison and ranking; model selection
  - Regression
  - M-estimation

Relevance and Applications

- **Mathematical / Statistical Applications:**
  - Forecast evaluation, comparison and ranking; model selection
  - Regression
  - M-estimation
- **Mathematical Finance**; banking; regulation:
  - Quantitative risk management;
  - Backtesting

Relevance and Applications

- **Mathematical / Statistical Applications:**
  - Forecast evaluation, comparison and ranking; model selection
  - Regression
  - M-estimation
- **Mathematical Finance**; banking; regulation:
  - Quantitative risk management;
  - Backtesting
- Economics; econometrics; business
- Meteorology
- Machine Learning
- Politics
- Sociology ($\rightsquigarrow$ 'Wisdom of the Crowds')

Regression

Classic situation: There is some parametric model $m \colon \Theta \times \mathbb{R} \to \mathbb{R}$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$Y = m_{\theta*}(X) + \varepsilon, \qquad \text{where} \quad \mathbf{E}[\varepsilon|X] = 0. \tag{1}$$

Regression

Classic situation: There is some parametric model $m \colon \Theta \times \mathbb{R} \to \mathbb{R}$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$Y = m_{\theta*}(X) + \varepsilon, \qquad \text{where} \quad \mathbf{E}[\varepsilon|X] = 0. \tag{1}$$

Equivalent form of (1):

$$\mathbf{E}[Y|X] = m_{\theta*}(X).$$

Regression

Classic situation: There is some parametric model $m\colon \Theta \times \mathbb{R} \to \mathbb{R}$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$Y = m_{\theta*}(X) + \varepsilon, \qquad \text{where} \quad \mathbf{E}[\varepsilon|X] = 0. \tag{1}$$

Equivalent form of (1):

$$\mathbf{E}[Y|X] = m_{\theta*}(X).$$

Find an estimator $\hat{\theta}_n$ for θ^* by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}(m_\theta(X_i) - Y_i)^2.$$

Regression

Classic situation: There is some parametric model $m \colon \Theta \times \mathbb{R} \to \mathbb{R}$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$Y = m_{\theta*}(X) + \varepsilon, \qquad \text{where} \quad \mathbf{E}[\varepsilon|X] = 0. \tag{1}$$

Equivalent form of (1):

$$\mathbf{E}[Y|X] = m_{\theta*}(X).$$

Find an estimator $\hat{\theta}_n$ for θ^* by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}(m_\theta(X_i) - Y_i)^2.$$

Relying in the fact that

$$\theta^* \in \arg\min_{\theta\in\Theta} \mathbf{E}(m_\theta(X) - Y)^2 \qquad \Big\{\theta^* \in \arg\min_{\theta\in\Theta} \mathbf{E}\big[(m_\theta(X) - Y)^2|X\big]\Big\}$$

Regression

Classic situation: There is some parametric model $m\colon \Theta \times \mathbb{R} \to \mathbb{R}$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$Y = m_{\theta*}(X) + \varepsilon, \qquad \text{where} \quad \mathbf{E}[\varepsilon|X] = 0. \tag{1}$$

Equivalent form of (1):

$$\mathbf{E}[Y|X] = m_{\theta*}(X).$$

Find an estimator $\hat{\theta}_n$ for θ^* by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n}(m_\theta(X_i) - Y_i)^2.$$

Relying in the fact that

$$\theta^* \in \arg\min_{\theta\in\Theta} \mathbf{E}(m_\theta(X) - Y)^2 \qquad \left\{\theta^* \in \arg\min_{\theta\in\Theta} \mathbf{E}\big[(m_\theta(X) - Y)^2|X\big]\right\}$$

However, instead of squared loss, we could use any strictly consistent scoring function for the mean functional.

Regression II

General situation: There is some parametric model $m\colon \Theta \times \mathbb{R}^\ell \to \mathbb{R}^k$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$T(\mathcal{L}(Y|X)) = m_{\theta*}(X).$$

Regression II

General situation: There is some parametric model $m\colon \Theta \times \mathbb{R}^\ell \to \mathbb{R}^k$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$T(\mathcal{L}(Y|\, X)) = m_{\theta*}(X).$$

Assume that $S\colon \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}$ is a strictly consistent scoring function for T.

Regression II

General situation: There is some parametric model $m\colon \Theta \times \mathbb{R}^\ell \to \mathbb{R}^k$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$T(\mathcal{L}(Y|\,X)) = m_{\theta*}(X).$$

Assume that $S\colon \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}$ is a strictly consistent scoring function for T.

Find an estimator $\hat{\theta}_n$ for θ^* by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} S(m_\theta(X_i), Y_i).$$

Regression II

General situation: There is some parametric model $m\colon \Theta \times \mathbb{R}^\ell \to \mathbb{R}^k$ and we assume that there is some true parameter $\theta^* \in \Theta$ such that

$$T(\mathcal{L}(Y|\,X)) = m_{\theta*}(X).$$

Assume that $S\colon \mathbb{R}^k \times \mathbb{R}^d \to \mathbb{R}$ is a strictly consistent scoring function for T.

Find an estimator $\hat{\theta}_n$ for θ^* by

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} \frac{1}{n}\sum_{i=1}^{n} S(m_\theta(X_i), Y_i).$$

Relying in the fact that

$$\theta^* \in \arg\min_{\theta\in\Theta} \mathbf{E}\big[S(m_\theta(X), Y)\big] \qquad \big\{\theta^* \in \arg\min_{\theta\in\Theta} \mathbf{E}\big[S(m_\theta(X), Y)|X\big]\big\}$$

The Elicitation Problem

Fix some functional $T \colon \mathcal{F} \to \mathsf{A}$.

(i) Is T elicitable?

(ii) What is the class of (strictly) consistent scoring functions for T?

(iii) What is a particularly good choice of a scoring function?

(iv) What to do if T is not elicitable?

[1] As shown in Fissler and Ziegel (2016, Annals of Statistics)

The Elicitation Problem

Fix some functional $T \colon \mathcal{F} \to \mathsf{A}$.

(i) Is T elicitable?

(ii) What is the class of (strictly) consistent scoring functions for T?

(iii) What is a particularly good choice of a scoring function?

(iv) What to do if T is not elicitable?

T	$S(x, y)$
mean	$(x-y)^2$
median	$\lvert x-y\rvert$
τ-expectile	$\lvert \mathbb{1}\{y \leqslant x\} - \tau\rvert (x-y)^2$
α-quantile	$\lvert \mathbb{1}\{y \leqslant x\} - \alpha\rvert\, \lvert x-y\rvert$

[1] As shown in Fissler and Ziegel (2016, Annals of Statistics)

The Elicitation Problem

Fix some functional $T \colon \mathcal{F} \to \mathsf{A}$.

(i) Is T elicitable?

(ii) What is the class of (strictly) consistent scoring functions for T?

(iii) What is a particularly good choice of a scoring function?

(iv) What to do if T is not elicitable?

T	$S(x, y)$
mean	$(x-y)^2$
median	$\lvert x-y \rvert$
τ-expectile	$\lvert \mathbb{1}\{y \leqslant x\} - \tau \rvert (x-y)^2$
α-quantile	$\lvert \mathbb{1}\{y \leqslant x\} - \alpha \rvert \, \lvert x-y \rvert$
variance	×
Expected Shortfall	×

[1] As shown in Fissler and Ziegel (2016, Annals of Statistics)

The Elicitation Problem

Fix some functional $T\colon \mathcal{F} \to \mathsf{A}$.

(i) Is T elicitable?

(ii) What is the class of (strictly) consistent scoring functions for T?

(iii) What is a particularly good choice of a scoring function?

(iv) What to do if T is not elicitable?

T	$S(x, y)$
mean	$(x-y)^2$
median	$\lvert x-y\rvert$
τ-expectile	$\lvert \mathbb{1}\{y \leqslant x\} - \tau\rvert (x-y)^2$
α-quantile	$\lvert \mathbb{1}\{y \leqslant x\} - \alpha\rvert \, \lvert x-y\rvert$
variance	×
Expected Shortfall	×
(mean, variance)	✓
(Value at Risk, Expected Shortfall)	✓[1]

[1] As shown in Fissler and Ziegel (2016, Annals of Statistics)

The Elicitation Problem

Fix some functional $T\colon \mathcal{F} \to \mathsf{A}$.

(i) Is T elicitable?

(ii) What is the class of (strictly) consistent scoring functions for T?

(iii) What is a particularly good choice of a scoring function?

(iv) What to do if T is not elicitable?

T	$S(x, y)$
mean	$(x-y)^2$
median	$\lvert x-y \rvert$
τ-expectile	$\lvert \mathbb{1}\{y \leqslant x\} - \tau \rvert (x-y)^2$
α-quantile	$\lvert \mathbb{1}\{y \leqslant x\} - \alpha \rvert \, \lvert x-y \rvert$
variance	×
Expected Shortfall	×
(mean, variance)	✓
(Value at Risk, Expected Shortfall)	✓[1]
identity (probabilistic forecast)	$S(F, y) = -\log(f(y))$

[1] As shown in Fissler and Ziegel (2016, Annals of Statistics)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).

$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting.
(Fissler, Ziegel and Gneiting, 2016; Risk)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting.
(Fissler, Ziegel and Gneiting, 2016; Risk)

(iv) Order-sensitivity, convexity and equivariance of scoring functions.
(Fissler and Ziegel, 2017)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting.
(Fissler, Ziegel and Gneiting, 2016; Risk)

(iv) Order-sensitivity, convexity and equivariance of scoring functions.
(Fissler and Ziegel, 2017)

(v) Semiparametric efficiency bounds for M- and Z-estimators of multidimensional functionals.
(Dimitriadis, Fissler and Ziegel, 2018+)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting.
(Fissler, Ziegel and Gneiting, 2016; Risk)

(iv) Order-sensitivity, convexity and equivariance of scoring functions.
(Fissler and Ziegel, 2017)

(v) Semiparametric efficiency bounds for M- and Z-estimators of multidimensional functionals.
(Dimitriadis, Fissler and Ziegel, 2018+)

(vi) Dichotomy concerning the elicitability of set-valued functionals.
(Fissler, Hlavinová and Rudloff, 2018+)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting.
(Fissler, Ziegel and Gneiting, 2016; Risk)

(iv) Order-sensitivity, convexity and equivariance of scoring functions.
(Fissler and Ziegel, 2017)

(v) Semiparametric efficiency bounds for M- and Z-estimators of multidimensional functionals.
(Dimitriadis, Fissler and Ziegel, 2018+)

(vi) Dichotomy concerning the elicitability of set-valued functionals.
(Fissler, Hlavinová and Rudloff, 2018+)

(vii) Elicitability and identifiability results for systemic risk measures.
(Fissler, Hlavinová and Rudloff, 2018+)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable. (Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting. (Fissler, Ziegel and Gneiting, 2016; Risk)

(iv) Order-sensitivity, convexity and equivariance of scoring functions. (Fissler and Ziegel, 2017)

(v) Semiparametric efficiency bounds for M- and Z-estimators of multidimensional functionals. (Dimitriadis, Fissler and Ziegel, 2018+)

(vi) Dichotomy concerning the elicitability of set-valued functionals. (Fissler, Hlavinová and Rudloff, 2018+)

(vii) Elicitability and identifiability results for systemic risk measures. (Fissler, Hlavinová and Rudloff, 2018+)

Main Achievements

Main contributions to the Elicitation Problem for vector-valued and set-valued functionals.

(i) Necessary characterization of strictly consistent scoring functions (Osband's Principle).
$\rightsquigarrow$ Multiple quantiles / expectiles. (Fissler and Ziegel, 2016; AoS)

(ii) Value at Risk and Expected Shortfall are jointly elicitable.
(Characterization of strictly consistent scoring functions also for spectral risk measures with finite support.) (Fissler and Ziegel, 2016; AoS)

(iii) Clarification of the role of elicitability for backtesting.
(Fissler, Ziegel and Gneiting, 2016; Risk)

(iv) Order-sensitivity, convexity and equivariance of scoring functions.
(Fissler and Ziegel, 2017)

(v) Semiparametric efficiency bounds for M- and Z-estimators of multidimensional functionals.
(Dimitriadis, Fissler and Ziegel, 2018+)

(vi) Dichotomy concerning the elicitability of set-valued functionals.
(Fissler, Hlavinová and Rudloff, 2018+)

(vii) Elicitability and identifiability results for systemic risk measures.
(Fissler, Hlavinová and Rudloff, 2018+)

Brief flavour of (i), (ii), (iii), and (vi).

One-dimensional functionals

Theorem 3 (Convex level sets, Osband, 1985)

Let $T\colon \mathcal{F} \to \mathsf{A}$ be an elicitable functional and $\mathcal{F}$ be convex. Then, for any $a \in \mathsf{A}$, the level sets

$$T^{-1}(\{a\}) \subseteq \mathcal{F}$$

are convex.

One-dimensional functionals

Theorem 3 (Convex level sets, Osband, 1985)

Let $T\colon \mathcal{F} \to \mathsf{A}$ be an elicitable functional and $\mathcal{F}$ be convex. Then, for any $a \in \mathsf{A}$, the level sets

$$T^{-1}(\{a\}) \subseteq \mathcal{F}$$

are convex.

Remarks:

- This shows that the variance or ES are generally not elicitable.

$$\mathrm{Var}(\delta_x) = \mathrm{Var}(\delta_y) = 0, \quad \mathrm{Var}\big(\lambda\delta_x + (1-\lambda)\delta_y\big) = \lambda(1-\lambda)(x-y)^2\,.$$

One-dimensional functionals

Theorem 3 (Convex level sets, Osband, 1985)

Let $T\colon \mathcal{F} \to \mathsf{A}$ be an elicitable functional and $\mathcal{F}$ be convex. Then, for any $a \in \mathsf{A}$, the level sets

$$T^{-1}(\{a\}) \subseteq \mathcal{F}$$

are convex.

Remarks:

- This shows that the variance or ES are generally not elicitable.

$$\mathrm{Var}(\delta_x) = \mathrm{Var}(\delta_y) = 0, \quad \mathrm{Var}\big(\lambda\delta_x + (1-\lambda)\delta_y\big) = \lambda(1-\lambda)(x-y)^2\,.$$

- Steinwart et al. (2014) showed that for $k = 1$ and under some regularity assumptions on T, cls are also sufficient for elicitability.

One-dimensional functionals

Theorem 3 (Convex level sets, Osband, 1985)

Let $T\colon \mathcal{F} \to \mathsf{A}$ be an elicitable functional and $\mathcal{F}$ be convex. Then, for any $a \in \mathsf{A}$, the level sets

$$T^{-1}(\{a\}) \subseteq \mathcal{F}$$

are convex.

Remarks:

- This shows that the variance or ES are generally not elicitable.

$$\mathrm{Var}(\delta_x) = \mathrm{Var}(\delta_y) = 0, \quad \mathrm{Var}\big(\lambda\delta_x + (1-\lambda)\delta_y\big) = \lambda(1-\lambda)(x-y)^2\,.$$

- Steinwart et al. (2014) showed that for $k = 1$ and under some regularity assumptions on T, cls are also sufficient for elicitability.
- This argument is independent of the dimension of T.

One-dimensional functionals

Theorem 3 (Convex level sets, Osband, 1985)

Let $T\colon \mathcal{F} \to \mathsf{A}$ be an elicitable functional and $\mathcal{F}$ be convex. Then, for any $a \in \mathsf{A}$, the level sets

$$T^{-1}(\{a\}) \subseteq \mathcal{F}$$

are convex.

Remarks:

- This shows that the variance or ES are generally not elicitable.

$$\mathrm{Var}(\delta_x) = \mathrm{Var}(\delta_y) = 0, \quad \mathrm{Var}\big(\lambda\delta_x + (1-\lambda)\delta_y\big) = \lambda(1-\lambda)(x-y)^2\,.$$

- Steinwart et al. (2014) showed that for $k = 1$ and under some regularity assumptions on T, cls are also sufficient for elicitability.
- This argument is independent of the dimension of T.
- For $k > 1$, it is an open question if cls are sufficient.

Vector-valued functionals

Lemma 4

If $T_1, \dots, T_k$ are elicitable, then the vector $(T_1, \dots, T_k)$ is elicitable.

Vector-valued functionals

Lemma 4

If $T_1, \ldots, T_k$ are elicitable, then the vector $(T_1, \ldots, T_k)$ is elicitable.

Theorem 5 (Revelation Principle, Osband, 1985)

If $T\colon \mathcal{F} \to \mathsf{A}$ is elicitable then any bijection $g \circ T\colon \mathcal{F} \to \mathsf{A}'$ is elicitable.

If $S(x, y)$ is strictly consistent for T, then $S(g^{-1}(x'), y)$ is strictly consistent for $g \circ T$.

Vector-valued functionals

Lemma 4

If $T_1, \ldots, T_k$ are elicitable, then the vector $(T_1, \ldots, T_k)$ is elicitable.

Theorem 5 (Revelation Principle, Osband, 1985)

If $T \colon \mathcal{F} \to \mathsf{A}$ is elicitable then any bijection $g \circ T \colon \mathcal{F} \to \mathsf{A}'$ is elicitable.

If $S(x, y)$ is strictly consistent for T, then $S(g^{-1}(x'), y)$ is strictly consistent for $g \circ T$.

$\rightsquigarrow$ The pair (mean, variance) is elicitable even though variance alone is not elicitable.

Vector-valued functionals

Lemma 4

If $T_1, \ldots, T_k$ are elicitable, then the vector $(T_1, \ldots, T_k)$ is elicitable.

Theorem 5 (Revelation Principle, Osband, 1985)

If $T\colon \mathcal{F} \to \mathrm{A}$ is elicitable then any bijection $g \circ T\colon \mathcal{F} \to \mathrm{A}'$ is elicitable.

If $S(x, y)$ is strictly consistent for T, then $S(g^{-1}(x'), y)$ is strictly consistent for $g \circ T$.

$\rightsquigarrow$ The pair (mean, variance) is elicitable even though variance alone is not elicitable.

Question

Are there elicitable functionals that are not a bijection of functionals with elicitable components only?

Value at Risk vs. Expected Shortfall

Value at Risk (VaR) and Expected Shortfall (ES) are the most commonly used risk measures in practice.

Definition 6

Let Y be an asset, $Y \sim F$, $\alpha \in (0,1)$. Then

$$\mathrm{VaR}_\alpha(F) := \inf\{x \in \mathbb{R} : F(x) \geqslant \alpha\},$$

$$\mathrm{ES}_\alpha(F) := \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\beta(F)\,\mathrm{d}\beta = \mathbf{E}_F\,[Y|\, Y \leqslant \mathrm{VaR}_\alpha(Y)].$$

- Profits amount to positive values of Y.
- We consider α close to zero (e.g. $\alpha = 0.01$, or $\alpha = 0.025$).
- Risky positions yield large negative values of VaR_α and ES_α.
  $\rightsquigarrow$ We work with utility functions instead of risk measures.

Value at Risk vs. Expected Shortfall (II)

Ongoing debate about the choice of a risk measure for regulatory purposes.

Value at Risk vs. Expected Shortfall (II)

Ongoing debate about the choice of a risk measure for regulatory purposes.

Properties of VaR_α as a risk measure:

$(+)$ It is elicitable, if the distributions in $\mathcal{F}$ have unique α-quantiles.

$(-)$ It is generally not superadditive (hence, not coherent).

$(-)$ It fails to take the size of losses beyond the level α into account.

Value at Risk vs. Expected Shortfall (II)

Ongoing debate about the choice of a risk measure for regulatory purposes.

Properties of VaR_α as a risk measure:

(+) It is elicitable, if the distributions in $\mathcal{F}$ have unique α-quantiles.

(−) It is generally not superadditive (hence, not coherent).

(−) It fails to take the size of losses beyond the level α into account.

Properties of ES_α as a risk measure:

(+) By definition, it considers the losses beyond the level α.

(+) It is superadditive (it is a coherent and comonotonically additive risk measure).

(−) It fails to have convex level sets and is consequently not elicitable; see Gneiting (2011).

Theorem 7 ((VaR, ES) – Fissler and Ziegel, AoS, 2016)

Let $\alpha \in (0,1)$*. Let* $\mathcal{F}$ *be a class of distribution functions on* $\mathbb{R}$ *with finite first moments. Let* $\mathsf{A}_0 = \{(x_1, x_2) \in \mathbb{R}^2 \colon x_1 \geqslant x_2\}$*, then any scoring function* $S \colon \mathsf{A}_0 \times \mathbb{R} \to \mathbb{R}$ *of the form*

$$\begin{aligned} S(x_1, x_2, y) = &\big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big) g(x_1) - \mathbb{1}\{y \leqslant x_1\} g(y) + a(y) \\ &+ \phi'(x_2)\Big(x_2 + \big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big)\frac{x_1}{\alpha} - \mathbb{1}\{y \leqslant x_1\}\frac{y}{\alpha}\Big) - \phi(x_2), \end{aligned} \tag{2}$$

is strictly $\mathcal{F}$*-consistent for* $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ *if*

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

Theorem 7 ((VaR, ES) – Fissler and Ziegel, AoS, 2016)

Let $\alpha \in (0,1)$. Let $\mathcal{F}$ be a class of distribution functions on $\mathbb{R}$ with finite first moments. Let $\mathsf{A}_0 = \{(x_1, x_2) \in \mathbb{R}^2 \colon x_1 \geqslant x_2\}$, then any scoring function $S \colon \mathsf{A}_0 \times \mathbb{R} \to \mathbb{R}$ of the form

$$\begin{aligned} S(x_1, x_2, y) = &\big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big) g(x_1) - \mathbb{1}\{y \leqslant x_1\} g(y) + a(y) \\ &+ \phi'(x_2)\Big(x_2 + \big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big)\frac{x_1}{\alpha} - \mathbb{1}\{y \leqslant x_1\}\frac{y}{\alpha}\Big) - \phi(x_2), \end{aligned} \tag{2}$$

is strictly $\mathcal{F}$-consistent for $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ if

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

Under mild regularity conditions, all strictly $\mathcal{F}$-consistent scoring functions for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ are of the form (2).

Theorem 7 ((VaR, ES) – Fissler and Ziegel, AoS, 2016)

Let $\alpha \in (0,1)$. Let $\mathcal{F}$ be a class of distribution functions on $\mathbb{R}$ with finite first moments. Let $\mathsf{A}_0 = \{(x_1, x_2) \in \mathbb{R}^2 \colon x_1 \geqslant x_2\}$, then any scoring function $S \colon \mathsf{A}_0 \times \mathbb{R} \to \mathbb{R}$ of the form

$$S(x_1, x_2, y) = \big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big) g(x_1) - \mathbb{1}\{y \leqslant x_1\} g(y) + a(y) \qquad (2)$$
$$+ \phi'(x_2)\Big(x_2 + \big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big)\frac{x_1}{\alpha} - \mathbb{1}\{y \leqslant x_1\}\frac{y}{\alpha}\Big) - \phi(x_2),$$

is strictly $\mathcal{F}$-consistent for $T = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ if

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

Under mild regularity conditions, all strictly $\mathcal{F}$-consistent scoring functions for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ are of the form (2).

$\rightsquigarrow$ Comparative Backtests of Diebold-Mariano type are possible; see Fissler, Ziegel and Gneiting (2016; Risk).

Identification function (moment function) $V\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$

$$\mathbf{E}_F[V(x, Y)] = 0 \quad \Longleftrightarrow \quad x = T(F)$$

Identification function (moment function) $V\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$

$$\mathbf{E}_F[V(x, Y)] = 0 \quad \iff \quad x = T(F)$$

T	$V(x, y)$
mean	$x - y$
α-quantile	$\mathbb{1}\{y \leqslant x\} - \alpha$

Identification function (moment function) $V\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$

$$\mathbf{E}_F[V(x, Y)] = 0 \quad \iff \quad x = T(F)$$

T	$V(x, y)$
mean	$x - y$
α-quantile	$\mathbb{1}\{y \leqslant x\} - \alpha$
$(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$	$\begin{pmatrix} \mathbb{1}\{y \leqslant x_1\} - \alpha \\ x_2 + \big(\mathbb{1}\{y \leqslant x_1\} - \alpha\big)x_1/\alpha - \mathbb{1}\{y \leqslant x_1\}y/\alpha \end{pmatrix}$

Identification function (moment function) $V\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$

$$\mathbf{E}_F[V(x, Y)] = 0 \quad \Longleftrightarrow \quad x = T(F)$$

T	$V(x, y)$
mean	$x - y$
α-quantile	$\mathbb{1}\{y \leqslant x\} - \alpha$
$(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$	$\begin{pmatrix} \mathbb{1}\{y \leqslant x_1\} - \alpha \\ x_2 + (\mathbb{1}\{y \leqslant x_1\} - \alpha)x_1/\alpha - \mathbb{1}\{y \leqslant x_1\}y/\alpha \end{pmatrix}$

Theorem 8 (Osband's Principle; Fissler and Ziegel, AoS; 2016))

Let $T\colon \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ be a surjective, elicitable and identifiable functional with a strict $\mathcal{F}$-identification function $V\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$.
Under some regularity assumptions, for any strictly $\mathcal{F}$-consistent scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ there exists a matrix-valued function $h\colon \mathrm{int}(\mathsf{A}) \to \mathbb{R}^{k\times k}$ such that

$$\nabla_x \mathbf{E}_F[S(x, Y)] = h(x)\, \mathbf{E}_F[V(x, Y)]$$

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

Second-order Osband's Principle

Theorem 9 (Osband's Principle; Fissler and Ziegel, AoS; 2016))

Let $T\colon \mathcal{F} \to \mathsf{A} \subseteq \mathbb{R}^k$ be a surjective, elicitable and identifiable functional with a strict $\mathcal{F}$-identification function $V\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}^k$.
Under some regularity assumptions, for any strictly $\mathcal{F}$-consistent scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ there exists a matrix-valued function $h\colon \mathrm{int}(\mathsf{A}) \to \mathbb{R}^{k\times k}$ such that

$$\nabla_x \mathbf{E}_F[S(x, Y)] = h(x)\, \mathbf{E}_F[V(x, Y)]$$

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

Second-order

Under some smoothness conditions, we can even exploit second order conditions: the Hessian

$$\nabla_x^2 \, \mathbf{E}_F\,[S(x, Y)] \in \mathbb{R}^{k\times k}$$

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

$\rightsquigarrow$ This gives a lot of information about the matrix $h(x)$.

Osband's Principle: Examples for $k = 1$

Proposition 10 (Gneiting, 2011)

(a) *Under some regularity conditions, $S\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a strictly consistent scoring function for the mean if and only if*

$$S(x, y) = \phi(y) - \phi(x) + \phi'(x)(x - y) + a(y)$$

$$\partial_x \mathbf{E}_F[S(x, Y)] = \phi''(x)(x - \mathbf{E}_F[Y]),$$

where $\phi\colon \mathbb{R} \to \mathbb{R}$ is strictly convex.

Osband's Principle: Examples for $k = 1$

Proposition 10 (Gneiting, 2011)

(a) *Under some regularity conditions, $S\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a strictly consistent scoring function for the mean if and only if*

$$S(x, y) = \phi(y) - \phi(x) + \phi'(x)(x - y) + a(y)$$

$$\partial_x \mathbf{E}_F [S(x, Y)] = \phi''(x)(x - \mathbf{E}_F [Y]),$$

where $\phi\colon \mathbb{R} \to \mathbb{R}$ is strictly convex.

(b) *Under some regularity conditions, $S\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a strictly consistent scoring function for the α-quantile, $\alpha \in (0, 1)$ if and only if*

$$S(x, y) = (\mathbb{1}\{y \leqslant x\} - \alpha)\big(g(x) - g(y)\big) + a(y)$$

$$\partial_x \mathbf{E}_F [S(x, Y)] = g'(x)(F(x) - \alpha),$$

where $g\colon \mathbb{R} \to \mathbb{R}$ is strictly increasing.

Relevance of Elicitability to Backtesting

Prediction-observation triples

$$(v_t, e_t, Y_t)_{t=1,\dots,N}$$

v_t: VaR_α prediction for time point t
e_t: ES_α prediction for time point t
Y_t: Realization at time point t

Traditional backtesting...

...aims at testing of the null hypothesis

H_0^C: "The risk measure estimates at hand are correct."

- Calculate some test statistic T_1 based on observations $(v_t, e_t, Y_t)_{t=1,\dots,N}$ such that we know the distribution of T_1 (approximately) under H_0^C.
- Backtesting decision: If we do not reject H_0^C, the risk measure estimates at hand are adequate.

Traditional backtesting...

...aims at testing of the null hypothesis

H_0^C: "The risk measure estimates at hand are correct."

- Calculate some test statistic T_1 based on observations $(v_t, e_t, Y_t)_{t=1,\dots,N}$ such that we know the distribution of T_1 (approximately) under H_0^C.
- Backtesting decision: If we do not reject H_0^C, the risk measure estimates at hand are adequate.

- Elicitability is not relevant.
- Does not respect increasing information sets.
- Does not give guidance for decision between methods.

Comparative backtesting

H_0^- : "The risk measure estimates at hand are *at least as good* as the ones from the standard procedure."

Comparative backtesting

H_0^- : "The risk measure estimates at hand are *at least as good* as the ones from the standard procedure."

- Internal model: $(v_t, e_t, Y_t)_{t=1,\dots,N} \rightsquigarrow \mathbf{S}_N = \frac{1}{N}\sum_{t=1}^N S(v_t, e_t, Y_t)$
- Standard model: $(v_t^*, e_t^*, Y_t)_{t=1,\dots,N} \rightsquigarrow \mathbf{S}_N^* = \frac{1}{N}\sum_{t=1}^N S(v_t^*, e_t^*, Y_t)$

(Asymptotically normal) test statistic:

$$T_2 = \frac{\mathbf{S}_N - \mathbf{S}_N^*}{\sigma_N},$$

where σ_N is a suitable estimate of the standard deviation.

Comparative backtesting

H_0^- : "The risk measure estimates at hand are *at least as good* as the ones from the standard procedure."

- Internal model: $(v_t, e_t, Y_t)_{t=1,\dots,N} \rightsquigarrow \mathbf{S}_N = \frac{1}{N}\sum_{t=1}^N S(v_t, e_t, Y_t)$
- Standard model: $(v_t^*, e_t^*, Y_t)_{t=1,\dots,N} \rightsquigarrow \mathbf{S}_N^* = \frac{1}{N}\sum_{t=1}^N S(v_t^*, e_t^*, Y_t)$

(Asymptotically normal) test statistic:

$$T_2 = \frac{\mathbf{S}_N - \mathbf{S}_N^*}{\sigma_N},$$

where σ_N is a suitable estimate of the standard deviation.

- Under H_0^-: Expectation of T_2 is $\leqslant 0$.
- Backtesting decision: If we do not reject H_0^-, the risk measure estimates at hand are acceptable (compared to the standard).

(Diebold and Mariano, 1995, Giacomini and White, 2006)

Some comments

- Elicitability is crucial.
- Allows for sensible comparison between methods.
- Respects increasing information sets (Holzmann and Eulert, 2014).

Some comments

- Elicitability is crucial.
- Allows for sensible comparison between methods.
- Respects increasing information sets (Holzmann and Eulert, 2014).
- H_0^C and H_0^- are anti-conservative: Passing the backtest does not imply the validity of the respective null hypothesis.

  "[...] the null hypothesis is never proved or established, but it is possibly disproved, in the course of experimentation." (Fisher, 1949)

Some comments

- Elicitability is crucial.
- Allows for sensible comparison between methods.
- Respects increasing information sets (Holzmann and Eulert, 2014).
- H_0^C and H_0^- are anti-conservative: Passing the backtest does not imply the validity of the respective null hypothesis.

  "[...] the null hypothesis is never proved or established, but it is possibly disproved, in the course of experimentation." (Fisher, 1949)
- We suggest a reversed onus of proof:
  Banks are obliged to demonstrate the superiority of the internal model.
  (Similar to regulatory practice in the health sector)

Conservative comparative backtesting

H_0^+ : "The risk measure estimates at hand are *at most as good* as the ones from the standard procedure."

- Internal model: $\mathbf{S}_N = \frac{1}{N}\sum_{t=1}^{N} S(v_t, e_t, Y_t)$
- Standard model: $\mathbf{S}_N^* = \frac{1}{N}\sum_{t=1}^{N} S(v_t^*, e_t^*, Y_t)$

(Asymptotically normal) test statistic:

$$T_2 = \frac{\mathbf{S}_N - \mathbf{S}_N^*}{\sigma_N},$$

where σ_N is a suitable estimate of the standard deviation.

- Under H_0^+: Expectation of T_2 is $\geqslant 0$.
- Backtesting decision: If we reject H_0^+, the risk measure estimates at hand are acceptable (compared to the standard).

(Diebold and Mariano, 1995, Giacomini and White, 2006)

Three zone approaches

BIS three zone approach for VaR_α

- Traditional backtest: One-sided binomial test.
- Backtesting decision:

	Red	Yellow	Green
p-value	very small	moderately small	sufficiently big

- Generalisation of three zone approach for ES_α by Costanzino and Curran (2015).

Three zone approaches

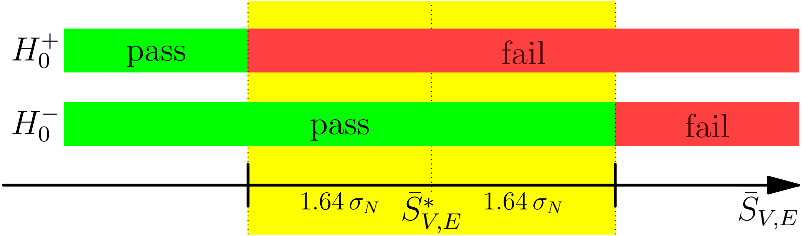
BIS three zone approach for VaR_α

- Traditional backtest: One-sided binomial test.
- Backtesting decision:

	Red	Yellow	Green
p-value	very small	moderately small	sufficiently big

- Generalisation of three zone approach for ES_α by Costanzino and Curran (2015).

Three zone approach for comparative backtesting



A numerical illustration

$(\mu_t)_{t=1,\dots,N}$ iid standard normal,

$$Y_t \sim \mathcal{N}(\mu_t, 1), \quad \text{conditional on } \mu_t.$$

Scenario A	
(v_t, e_t)	$= (\mathrm{VaR}_\alpha(\mathcal{N}(\mu_t, 1)), \mathrm{ES}_\alpha(\mathcal{N}(\mu_t, 1)))$
(v_t^*, e_t^*)	$= (\mathrm{VaR}_\alpha(\mathcal{N}(0, 2)), \mathrm{ES}_\alpha(\mathcal{N}(0, 2)))$

The internal model is more informative,
hence superior to the standard model.

A numerical illustration

$(\mu_t)_{t=1,\dots,N}$ iid standard normal,

$$Y_t \sim \mathcal{N}(\mu_t, 1), \quad \text{conditional on } \mu_t.$$

Scenario A	
(v_t, e_t)	$= (\mathrm{VaR}_\alpha(\mathcal{N}(\mu_t, 1)), \mathrm{ES}_\alpha(\mathcal{N}(\mu_t, 1)))$
(v_t^*, e_t^*)	$= (\mathrm{VaR}_\alpha(\mathcal{N}(0, 2)), \mathrm{ES}_\alpha(\mathcal{N}(0, 2)))$

The internal model is more informative,
hence superior to the standard model.

Scenario B	
(v_t, e_t)	$= (\mathrm{VaR}_\alpha(\mathcal{N}(0, 2)), \mathrm{ES}_\alpha(\mathcal{N}(0, 2)))$
(v_t^*, e_t^*)	$= (\mathrm{VaR}_\alpha(\mathcal{N}(\mu_t, 1)), \mathrm{ES}_\alpha(\mathcal{N}(\mu_t, 1)))$

The standard model is more informative,
hence superior to the internal model.

A numerical illustration – cont'd

$N = 250$; 10'000 simulations

Scenario A		**Green**	Yellow	Red
Traditional	$\mathrm{VaR}_{0.01}$	89.35	10.65	0.00
Traditional	$\mathrm{ES}_{0.025}$	93.62	6.36	0.02
Comparative	$\mathrm{VaR}_{0.01}$	88.23	11.77	0.00
Comparative	$(\mathrm{VaR}_{0.025}, \mathrm{ES}_{0.025})$	87.22	12.78	0.00

A numerical illustration – cont'd

$N = 250$; 10'000 simulations

Scenario A		**Green**	Yellow	Red
Traditional	$\mathrm{VaR}_{0.01}$	89.35	10.65	0.00
Traditional	$\mathrm{ES}_{0.025}$	93.62	6.36	0.02
Comparative	$\mathrm{VaR}_{0.01}$	88.23	11.77	0.00
Comparative	$(\mathrm{VaR}_{0.025}, \mathrm{ES}_{0.025})$	87.22	12.78	0.00

Scenario B		Green	Yellow	**Red**
Traditional	$\mathrm{VaR}_{0.01}$	89.33	10.67	0.00
Traditional	$\mathrm{ES}_{0.025}$	93.80	6.18	0.02
Comparative	$\mathrm{VaR}_{0.01}$	0.00	11.77	88.23
Comparative	$(\mathrm{VaR}_{0.025}, \mathrm{ES}_{0.025})$	0.00	12.78	87.22

Summary

- Elicitability is not relevant for traditional backtesting.
- Elicitability is useful for model selection, estimation, forecast comparison and ranking.
- Comparative backtesting relies on elicitability, using H_0^+ it is conservative in nature and gives (more) incentive to improve predictions.

Set-valued functionals

- **Quantiles**

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

Set-valued functionals

- **Quantiles**

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- **Class of confidence intervals**

$$\mathcal{I}_\alpha(F) = \{(a, b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

Set-valued functionals

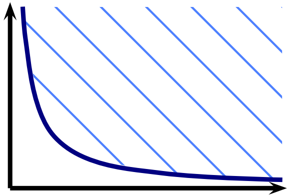
- **Quantiles**

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- **Class of confidence intervals**

$$\mathcal{I}_\alpha(F) = \{(a, b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

- **Systemic risk measures (Feinstein, Rudloff, Weber, 2017)**



$$R(Y) = \{k \in \mathbb{R}^n \mid \rho(\Lambda(Y + k)) \leqslant 0\}.$$

Set-valued functionals

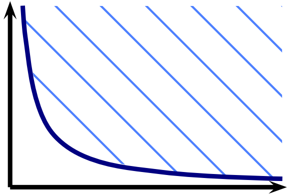
- **Quantiles**

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- **Class of confidence intervals**

$$\mathcal{I}_\alpha(F) = \{(a, b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

- **Systemic risk measures (Feinstein, Rudloff, Weber, 2017)**



$$R(Y) = \{k \in \mathbb{R}^n \mid \rho(\Lambda(Y + k)) \leqslant 0\}.$$

- **Further spatial examples:** Area of flood, disease, landfall of a hurricane etc.

Two modes of elicitability

- Example of the α-quantile

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

Two modes of elicitability

- Example of the α-quantile

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- Choice of the action domain A:

Two modes of elicitability

- Example of the α-quantile

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- Choice of the action domain A:

  $\mathsf{A} = \mathbb{R}$: The forecasts are points in $\mathbb{R}$. There are multiple best actions, namely every $x \in q_\alpha(F)$.
  $\rightsquigarrow$ The functional T is set-valued, that is

$$T \colon \mathcal{F} \to 2^{\mathsf{A}}.$$

Two modes of elicitability

- Example of the α-quantile

$$q_\alpha(F) = \{x \in \mathbb{R} \mid \lim_{t \uparrow x} F(t) \leqslant \alpha \leqslant F(x)\} \subset \mathbb{R}.$$

- Choice of the action domain A:

$\mathsf{A} = \mathbb{R}$: The forecasts are points in $\mathbb{R}$. There are multiple best actions, namely every $x \in q_\alpha(F)$.
$\rightsquigarrow$ The functional T is set-valued, that is

$$T\colon \mathcal{F} \to 2^{\mathsf{A}}.$$

$\mathsf{A} \subseteq 2^{\mathbb{R}}$: The forecasts are subsets of $\mathbb{R}$. These are points in the power set $\mathsf{A} \subseteq 2^{\mathbb{R}}$. There is a unique best action namely $x = q_\alpha(F)$.
$\rightsquigarrow$ The functional T is point-valued in some space $\mathsf{A} \subseteq 2^{\mathbb{R}}$, that is,

$$T\colon \mathcal{F} \to \mathsf{A}.$$

Two modes of elicitability

Definition 11

(a) A functional $T\colon \mathcal{F} \to 2^{\mathsf{A}}$ is selectively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(t, Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $t \in T(F)$ and for all $x \in \mathsf{A} \setminus T(F)$.

Two modes of elicitability

Definition 11

(a) A functional $T\colon \mathcal{F} \to 2^{\mathsf{A}}$ is selectively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(t, Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $t \in T(F)$ and for all $x \in \mathsf{A} \setminus T(F)$.

(b) A functional $T\colon \mathcal{F} \to \mathsf{A}$ is exhaustively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(T(F), Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$, $x \neq T(F)$.

Two modes of elicitability

Definition 11

(a) A functional $T\colon \mathcal{F} \to 2^{\mathsf{A}}$ is selectively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(t, Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $t \in T(F)$ and for all $x \in \mathsf{A} \setminus T(F)$.

(b) A functional $T\colon \mathcal{F} \to \mathsf{A}$ is exhaustively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(T(F), Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$, $x \neq T(F)$.

Two modes of elicitability

Definition 11

(a) A functional $T\colon \mathcal{F} \to 2^{\mathsf{A}}$ is selectively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(t, Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $t \in T(F)$ and for all $x \in \mathsf{A} \setminus T(F)$.

(b) A functional $T\colon \mathcal{F} \to \mathsf{A}$ is exhaustively elicitable if there is a scoring function $S\colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(T(F), Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$, $x \neq T(F)$.

- For single-valued functionals such as the mean, the notions of selective and exhaustive elicitability are equivalent.

Two modes of elicitability

Definition 11

(a) A functional $T \colon \mathcal{F} \to 2^{\mathsf{A}}$ is selectively elicitable if there is a scoring function $S \colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(t, Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $t \in T(F)$ and for all $x \in \mathsf{A} \setminus T(F)$.

(b) A functional $T \colon \mathcal{F} \to \mathsf{A}$ is exhaustively elicitable if there is a scoring function $S \colon \mathsf{A} \times \mathsf{O} \to \mathbb{R}$ such that

$$\mathbf{E}_F[S(T(F), Y)] < \mathbf{E}_F[S(x, Y)]$$

for all $F \in \mathcal{F}$ and for all $x \in \mathsf{A}$, $x \neq T(F)$.

- For single-valued functionals such as the mean, the notions of selective and exhaustive elicitability are equivalent.
- Forecasting / regression in the exhaustive sense is more ambitious than in the selective sense!

Mutual exclusivity

Theorem 12 (Fissler, Hlavinová, Rudloff (2018+))

Under weak regularity conditions, a set-valued functional is

- *either selectively elicitable*
- *or exhaustively elicitable*
- *or not elicitable at all.*

Mutual exclusivity

Theorem 12 (Fissler, Hlavinová, Rudloff (2018+))

Under weak regularity conditions, a set-valued functional is

- *either selectively elicitable*
- *or exhaustively elicitable*
- *or not elicitable at all.*

Novel structural insight of its own!

Mutual exclusivity

Theorem 12 (Fissler, Hlavinová, Rudloff (2018+))

Under weak regularity conditions, a set-valued functional is

- *either selectively elicitable*
- *or exhaustively elicitable*
- *or not elicitable at all.*

Novel structural insight of its own!

Results and implications:

- Quantiles are selectively elicitable, but not exhaustively elicitable!

$$S_\alpha(x, y) = \left(\mathbb{1}\{y \leqslant x\} - \alpha\right)\left(g(x) - g(y)\right).$$

Mutual exclusivity

Theorem 12 (Fissler, Hlavinová, Rudloff (2018+))

Under weak regularity conditions, a set-valued functional is

- *either selectively elicitable*
- *or exhaustively elicitable*
- *or not elicitable at all.*

Novel structural insight of its own!

Results and implications:

- Quantiles are selectively elicitable, but not exhaustively elicitable!

$$S_\alpha(x, y) = \big(\mathbb{1}\{y \leqslant x\} - \alpha\big)\big(g(x) - g(y)\big)\,.$$

- Many systemic risk measures are exhaustively elicitable, but not selectively elicitable.

$$S_R(K, y) = -\int_K V_\rho\big(0, \Lambda(y + k)\big)\,\pi(\mathrm{d}k)\,.$$

Mutual exclusivity

Theorem 12 (Fissler, Hlavinová, Rudloff (2018+))

Under weak regularity conditions, a set-valued functional is

- *either selectively elicitable*
- *or exhaustively elicitable*
- *or not elicitable at all.*

Novel structural insight of its own!

Results and implications:

- Quantiles are selectively elicitable, but not exhaustively elicitable!

$$S_\alpha(x,y) = \big(\mathbb{1}\{y \leqslant x\} - \alpha\big)\big(g(x) - g(y)\big)\,.$$

- Many systemic risk measures are exhaustively elicitable, but not selectively elicitable.

$$S_R(K,y) = -\int_K V_\rho\big(0, \Lambda(y+k)\big)\,\pi(\mathrm{d}k)\,.$$

- Confidence intervals.

Confidence Intervals

Reminder:

$$\mathcal{I}_\alpha(F) = \{(a, b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

- $\mathcal{I}_\alpha$ is selectively identifiable with $V(a, b, y) = \mathbb{1}\{y \in (a, b]\} - \alpha$.

Confidence Intervals

Reminder:

$$\mathcal{I}_\alpha(F) = \{(a, b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

- $\mathcal{I}_\alpha$ is selectively identifiable with $V(a, b, y) = \mathbb{1}\{y \in (a, b]\} - \alpha$.
- It is presumably exhaustively elicitable.

Confidence Intervals

Reminder:

$$\mathcal{I}_\alpha(F) = \{(a,b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

- $\mathcal{I}_\alpha$ is selectively identifiable with $V(a,b,y) = \mathbb{1}\{y \in (a,b]\} - \alpha$.
- It is presumably exhaustively elicitable.
- It is not selectively elicitable!

Confidence Intervals

Reminder:

$$\mathcal{I}_\alpha(F) = \{(a, b) \in \mathbb{R}^2 \mid F(b) - F(a) = \alpha\}.$$

- $\mathcal{I}_\alpha$ is selectively identifiable with $V(a, b, y) = \mathbb{1}\{y \in (a, b]\} - \alpha$.
- It is presumably exhaustively elicitable.
- It is not selectively elicitable!
- One needs to have additional properties for selective elicitability:
  - Specify the endpoints as quantiles.
  - Take a 'symmetric' interval.
  - Shortest confidence interval does not work.
  - Centring around the median or mean also fails.

Further Reading

- Good introduction to elicitability:

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

- Elicitability of vector-valued functionals and elicitability of (VaR, ES):

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

- Backtesting and elicitability: T. Fissler, J. F. Ziegel, and T. Gneiting. Expected shortfall is jointly elicitable with value-at-risk: implications for backtesting.
  Risk Magazine, pages 58–61, January 2016

  N. Nolde and J. F. Ziegel. Elicitability and backtesting: Perspectives for banking regulation.
  Annals of Applied Statistics, 11(4):1833–1874, 12 2017

Further Reading II

- Secondary quality criteria:

T. Fissler and J. F. Ziegel. Order-sensitivity and equivariance of scoring functions. *Preprint*, 2017

T. Fissler and J. F. Ziegel. Convex and quasi-convex scoring functions. *In preparation*, 2018

- Measures of Systemic Risk:

Z. Feinstein, B. Rudloff, and S. Weber. Measures of Systemic Risk. *SIAMJ. Financial Math.*, 8:672–708, 2017

T. Fissler, J. Hlavinová, and B. Rudloff. Elicitability and identifiability of systemic risk measures. *In preparation*, 2018

Thank you for your attention!

Looking forward to our discussion!