E-backtesting

Ruodu Wang

http://sas.uwaterloo.ca/~wang

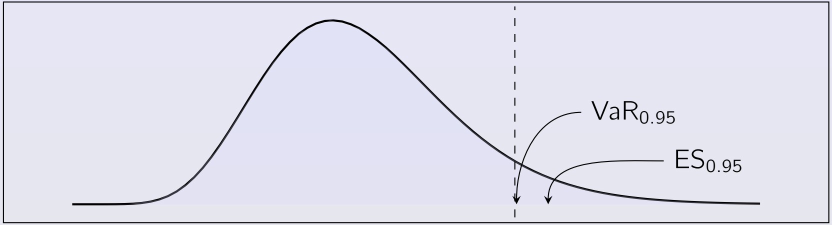
Department of Statistics and Actuarial Science
University of Waterloo

Finance Innovations & AI Seminar, Canadian Imperial Bank of Commerce
November 2023 (online)

Agenda

based on joint work with Qiuqi Wang (Georgia State) and Johanna Ziegel (Bern)

VaR and ES



Value-at-Risk (VaR), $p \in (0, 1)$

$\mathrm{VaR}_p : L^0 \to \mathbb{R}$,

$$\begin{aligned}\mathrm{VaR}_p(X) &= q_p(X) \\ &= \inf\{x \in \mathbb{R} : \mathbb{P}(X \leqslant x) \geqslant p\}\end{aligned}$$

(left-quantile)

Expected Shortfall (ES), $p \in (0, 1)$

$\mathrm{ES}_p : L^1 \to \mathbb{R}$,

$$\mathrm{ES}_p(X) = \frac{1}{1-p}\int_p^1 \mathrm{VaR}_q(X)\mathrm{d}q$$

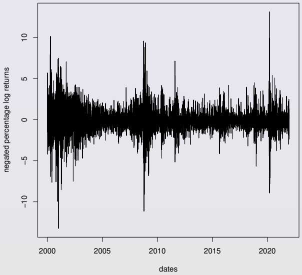
(also: TVaR/CVaR/AVaR)

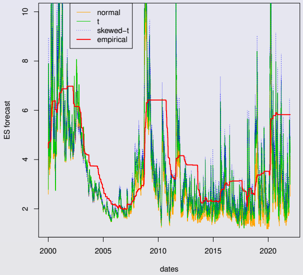
Some recent work on VaR vs ES

- Axiomatic characterizations
  - VaR: Kou/Peng' 16 OR; He/Peng' 18 OR; Liu/W.'21 MOR
  - ES: W./Zitikis'21 MS; Embrechts/Mao/Wang/W.'21 MF
- Risk sharing
  - Embrechts/Liu/W.'18 OR; Embrechts/Liu/Mao/W.'20 MP
- Robustness, optimization, calibration
  - Emberchts/Schied/W.'22 OR; Li/W.'23 JE
- Forecasting and backtesting ES
  - Fissler/Ziegel'16 AOS; Nolde/Ziegel'17 AOAS; Du/Escanciano'17 MS; Moldenhauer/Pitera'17 JRisk; Banulescu-Radu/Hurlin/Leymarie/Scaillet'21 MS; Bayer/Dimitriadis'22 JFEC; Hoga/Demetrescu'22 MS

Features in backtesting risk measures I

After each period (e.g., one trading day), a new observation comes in, and a risk forecast is announced by a financial institution (bank). Hypothesis testing methods are designed to assess the risk forecasts.





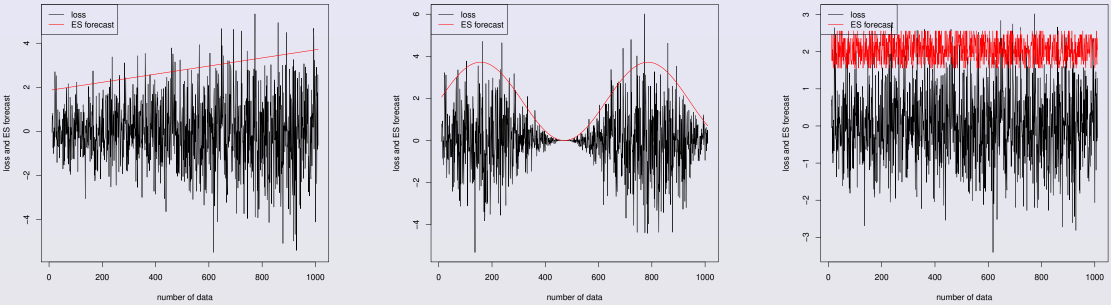
- Negated log-returns (in %) of the NASDAQ Composite index from Jan 16, 1996 to Dec 31, 2021
- Fitted (AR(1)-GARCH(1, 1)) or empirical $ES_{0.975}$ forecasts with moving window of 500

Features in backtesting risk measures II

Challenges

- The risks are neither independent nor identically distributed
- The predictions are even less clearly structured
- The regulator does not necessarily know or trust the underlying model used by the bank to produce a risk prediction
  - Only limited information is supplied by the bank, and the bank may make mistakes on models it provide.

Some examples



The firm is expanding the business | The firm has a business cycle | The firm has a poor forecast quality

Question: How do we evaluate the ES forecasts (the true ES is not known)?

Features in backtesting risk measures III

The regulator is concerned about underestimation of the risk measure, whereas overestimation (being conservative) is less of a concern.

Example from Lehman Brothers (2008)

- underestimated default/credit risks
- hid the true their leverage and debt

Example from Silicon Valley Bank (2023)

- underestimated interest rate risk
- basically no risk management

Backtesting risk measures

- Risk measure ρ to backtest
- Define

$$\mathcal{F}_{t-1} := \sigma(L_s : s \leqslant t-1)$$

- Daily observations
  - risk measure forecast r_t for $\rho(L_t)$ given $\mathcal{F}_{t-1}$
  - realized loss L_t

Hypothesis to test

$$H_0 : \begin{array}{c}\text{conditional on } \mathcal{F}_{t-1}: \\ r_t \geqslant \rho(L_t|\mathcal{F}_{t-1})\end{array} \quad \text{for } t = 1, \dots, T$$

Backtesting VaR

- Daily prediction $r_t = \widehat{\mathrm{VaR}}_p(L_t|\mathcal{F}_{t-1})$
- Daily realization L_t

Backtesting for fixed T

- Under H_0: $Y_t = \mathbb{1}_{\{L_t > r_t\}}$ are independent Bernoulli sample with mean at most $1 - p$
- $S_T = \sum_{t=1}^{T} Y_t \leqslant_{\mathrm{st}} \mathrm{Binomial}(T, 1-p)$
- Easy to construct p-values (reject if S_t large enough)
- Completely model free

Such a simple procedure does not exist for ES!

Backtesting risk measures

Objective: build up a backtesting method that

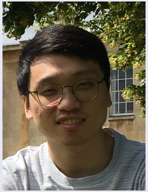
- is model free
- is anytime valid
- works for both ES and VaR

Progress

E-values

Vladimir Vovk
(Royal Holloway)

Aaditya Ramdas
(Carnegie Mellon)



Qiuqi Wang
(Georgia State)

Johanna Ziegel
(Bern)

- Vovk/W., E-values: Calibration, combination, and applications. Annals of Statistics, 2021
- W./Ramdas, False discovery rate control with e-values. JRSSB, 2022
- Grünwald/de Heide/Koolen, Safe testing. JRSSB, 2023
- Vovk/W., Confidence and discoveries with e-values. Statistical Science, 2023
- Waudby-Smith/Ramdas, Estimating means of bounded random variables by betting. JRSSB, 2023
- Wang/W./Ziegel, E-backtesting. Working paper, 2022, arXiv:2209.00991

What is an e-value?

- A hypothesis $\mathcal{H}$: a set of probability measures

Definition (E-variables, e-values, and e-processes)

(1) An e-variable for testing $\mathcal{H}$ is a non-negative random variable $E : \Omega \to [0, \infty]$ that satisfies $\int E \,\mathrm{d}H \leqslant 1$ for all $H \in \mathcal{H}$.

- Realized values of e-variables are e-values.

(2) Given a filtration, an e-process for testing $\mathcal{H}$ is a non-negative process $(E_t)_{t=0,1,\dots,n}$ such that $\int E_\tau \mathrm{d}H \leqslant 1$ for all stopping times τ and all $H \in \mathcal{H}$.

- For simple hypothesis $\{\mathbb{P}\}$
  - e-variable: non-negative random variable with mean $\leqslant 1$
  - e-process: (e.g.) non-negative supermartingale with initial value $\leqslant 1$

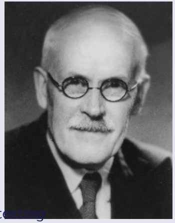
What is an e-value?

- E-test: $e(\text{data})$ large $\Longleftrightarrow$ reject $\mathcal{H}$
- P-test: $p(\text{data})$ small $\Longleftrightarrow$ reject $\mathcal{H}$
- E stands for expectation; P stands for probability
- Bayes factors (simple hypothesis) and likelihood ratios:

$$e(\text{data}) = \frac{\Pr(\text{data} \mid \mathbb{Q})}{\Pr(\text{data} \mid \mathbb{P})}$$

Sir Jeffreys

"*Users of these tests speak of the 5 per cent. point* [p-value of 5%] *in much the same way as I should speak of the* $K = 10^{-1/2}$ *point* [e-value of $10^{1/2}$], *and of the 1 per cent. point* [p-value of 1%] *as I should speak of the* $K = 10^{-1}$ *point* [e-value of 10]." (Theory of Probability, p.435, 3rd Ed., 1961)



Robustness advantages

- Validity for arbitrary dependence
  - robust to dependence
- They are easy to combine
  - robust to operations
- Flexible with regards to stop/continue procedures
  - robust to sampling and optimizing algorithms
- Non-asymptotic and model free
  - robust to model misspecification

Example in testing multiple hypotheses

Multi-armed bandit problems

Xu-W.-Ramdas'21 NeurIPS

- K arms
- null hypothesis k: arm k has mean reward at most 1
- strategy (k_t): at time $t \geqslant 1$, pull arm k_t, obtain an iid reward $X_{k_t,t} \geqslant 0$
- aim: quickly detect arms with mean > 1
  - or maximize profit, minimize regret, etc ...
- running reward: $M_{k,t} = \prod_{j=1}^{t} X_{k,j} \mathbb{1}_{\{k_j = k\}}$
- complicated dependence due to exploration/exploitation
- $M_{1,\tau}, \dots, M_{K,\tau}$ are e-values for any stopping time τ

Combining sequential e-values

Sequential e-variables

- $\mathbb{E}[X_t \mid X_1, \ldots, X_{t-1}] \leqslant 1$ for all $t = 1, 2, \ldots$

The general protocol

- Obtain sequential e-values $X_1, \ldots, X_t, \ldots$
- Decide predictable $\lambda_1, \ldots, \lambda_t, \cdots \in [0, 1]$
- Compute the e-process ($E_0 = 1$)

$$E_t = E_{t-1}(1 - \lambda_t + \lambda_t X_t) = \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s X_s)$$

- The only optimal procedures in the sense of Pareto or Wald   Vovk/W.'22

Progress

Risk forecasts and backtests

E-values

Model-free e-statistics

E-backtesting

Simulation and data analysis

Model-free e-statistics

Setting

- The model space $\mathcal{M}$ is a set of distributions on $\mathbb{R}$
- ρ is the risk measure to be tested
  - treated as a mapping on either $\mathcal{M}$ or $\mathcal{X}$
- $\phi : \mathcal{M} \to \mathbb{R}$ represents auxiliary statistics
- $\psi = (\rho, \phi)$ represents the collection of available statistical information

Remarks.

- If ϕ is a constant (we can take $\phi = 0$), then only the predicted value of ρ is used
- We omit ϕ if it is a constant
- ϕ may be d-dimensional in general, but we focus on $d = 0$ (constant ϕ) or $d = 1$

Model-free e-statistics

Intuitive thoughts in case ϕ is omitted

- $e(X, r)$ is an e-variable if $r \geqslant \rho(F_X)$ $\longleftarrow$ validity
- If r is under-specified, then $\mathbb{E}[e(X, r)] > 1$ $\longleftarrow$ consistency
- $r \mapsto e(X, r)$ should be decreasing $\longleftarrow$ monotonicity

Model-free e-statistics

Intuitive thoughts in case ϕ is omitted

- $e(X, r)$ is an e-variable if $r \geqslant \rho(F_X)$ $\longleftrightarrow$ validity
- If r is under-specified, then $\mathbb{E}[e(X, r)] > 1$ $\longleftrightarrow$ consistency
- $r \mapsto e(X, r)$ should be decreasing $\longleftrightarrow$ monotonicity

Definition (Model-free e-statistics)

A model-free e-statistic for $(\rho, \phi) : \mathcal{M} \to \mathbb{R}^2$ is a measurable function $e : \mathbb{R}^3 \to [0, \infty]$ satisfying $\int e(x, \rho(F), \phi(F))\mathrm{d}F \leqslant 1$ for each $F \in \mathcal{M}$. Moreover,

- e is testing ρ if $\int e(x, r, z)\mathrm{d}F(x) > 1$ for all $F \in \mathcal{M}$ with $\rho(F) > r$ and all (r, z) in the range of (ρ, ϕ).
- The test is strict if $r \mapsto e(x, r, z)$ is decreasing.

- We write $e(x, r) = e(x, r, 0)$ if $\phi = 0$ (so z does not matter)

Model-free e-statistics

Examples (convention: $0/0 = 1$ and $1/0 = \infty$)

- Let $\mathcal{M}$ be the set of distributions on $\mathbb{R}_+$ with a finite mean. The function $e(x, r) = x/r$ for $x, r \geqslant 0$ is a model-free e-statistic strictly testing the mean.

Model-free e-statistics

Examples (convention: $0/0 = 1$ and $1/0 = \infty$)

- Let $\mathcal{M}$ be the set of distributions on $\mathbb{R}_+$ with a finite mean. The function $e(x, r) = x/r$ for $x, r \geqslant 0$ is a model-free e-statistic strictly testing the mean.
- Let $\mathcal{M}$ be the set of distributions on $\mathbb{R}$ with a finite variance. The function $e(x, r, z) = (x - z)^2/r$ for $x, z \in \mathbb{R}$ and $r \geqslant 0$ is a model-free e-statistic for $(\mathbb{E}, \mathrm{Var})$ strictly testing the variance.

Model-free e-statistics

Examples (convention: $0/0 = 1$ and $1/0 = \infty$)

- Let $\mathcal{M}$ be the set of distributions on $\mathbb{R}_+$ with a finite mean. The function $e(x, r) = x/r$ for $x, r \geqslant 0$ is a model-free e-statistic strictly testing the mean.
- Let $\mathcal{M}$ be the set of distributions on $\mathbb{R}$ with a finite variance. The function $e(x, r, z) = (x - z)^2/r$ for $x, z \in \mathbb{R}$ and $r \geqslant 0$ is a model-free e-statistic for $(\mathbb{E}, \mathrm{Var})$ strictly testing the variance.
- Let $\mathcal{M}$ be the set of all distributions on $\mathbb{R}$ and take $p \in (0, 1)$. The function $e(x, r) = \mathbb{1}_{\{x>r\}}/(1 - p)$ for $x, r \in \mathbb{R}$ is a model-free e-statistic strictly testing VaR_p.

Model-free e-statistics for VaR

The model-free e-statistic strictly testing VaR_p

$$e_p^Q(x,r) = \frac{\mathbb{1}_{\{x>r\}}}{1-p}, \quad x,r \in \mathbb{R}$$

Theorem 1

For $p \in (0,1)$, all model-free e-statistics testing VaR_p that are continuous except at $x = r$ have the form

$$e'(x,r) = 1 - \lambda(r) + \lambda(r) e_p^Q(x,r), \quad x,r \in \mathbb{R},$$

for some continuous $\lambda : \mathbb{R} \to (0,1]$. Moreover, e' is strictly testing VaR_p if and only if $\lambda(r)$ is constant in r.

Model-free e-statistics for ES

Proposition 1

There does not exist a model-free e-statistic testing ES_p using solely the information of ES_p.

Define the function

$$e_p^{\mathrm{ES}}(x, r, z) = \frac{(x - z)_+}{(1 - p)(r - z)}, \quad x \in \mathbb{R}, \ z \leqslant r,$$

Theorem 2

For $p \in (0, 1)$, e_p^{ES} is a model-free e-statistic for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ strictly testing ES_p.

Model-free e-statistics for ES

Sketch of Proof. VaR-ES optimization formula: Rockafellar/Uryasev'02 JBF

$$\mathrm{VaR}_p(X) \in \arg\min_{x\in\mathbb{R}} \left\{ x + \frac{1}{1-p}\mathbb{E}[(X-x)_+] \right\}$$

$$\begin{aligned}\mathrm{ES}_p(X) &= \min_{x\in\mathbb{R}} \left\{ x + \frac{1}{1-p}\mathbb{E}[(X-x)_+] \right\} \\ &= \mathrm{VaR}_p(X) + \frac{1}{1-p}\mathbb{E}[(X-\mathrm{VaR}_p(X))_+]\end{aligned}$$

Hence,

$$\mathbb{E}[e_p^{\mathrm{ES}}(X, \mathrm{ES}_p(X), \mathrm{VaR}_p(X))] = \frac{\mathbb{E}[(X-\mathrm{VaR}_p(X))_+]}{(1-p)(\mathrm{ES}_p(X)-\mathrm{VaR}_p(X))} = 1$$

and

$$\mathbb{E}[e_p^{\mathrm{ES}}(X, r, z)] > 1 \quad \text{for any } z \leqslant r < \mathrm{ES}_p(X)$$

Model-free e-statistics for ES

Theorem 3

All model-free e-statistics for $(\mathrm{ES}_p, \mathrm{VaR}_p)$ *testing* ES_p *have the form*

$$e'(x,r,z) = 1 - \lambda(r,z) + \lambda(r,z) e_p^{\mathrm{ES}}(x,r,z), \quad x \in \mathbb{R},\ z \leqslant r.$$

for some $\lambda : \mathbb{R}^2 \to (0,1]$. *Moreover,* e' *is strictly testing* ES_p *if and only if both* $r \mapsto \lambda(r,z)$ *and* $r \mapsto (r-z)/\lambda(r,z)$ *are increasing.*

Progress

Risk forecasts and backtests

E-values

Model-free e-statistics

E-backtesting

Simulation and data analysis

E-backtesting

Daily observations

- ES forecast r_t
- VaR forecast z_t
- realized loss L_t

Hypothesis to test

$$H_0: \begin{array}{c}\text{conditional on } \mathcal{F}_{t-1}: \\ r_t \geqslant \mathrm{ES}_p(L_t|\mathcal{F}_{t-1}) \text{ and } z_t = \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1})\end{array} \quad \text{for } t = 1, \dots, T$$

A weaker hypothesis

$$H_0': \begin{array}{c}\text{conditional on } \mathcal{F}_{t-1}: \\ r_t - z_t \geqslant \mathrm{ES}_p(L_t|\mathcal{F}_{t-1}) - \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1}) \quad \text{for } t = 1, \dots, T \\ \text{and } z_t \geqslant \mathrm{VaR}_p(L_t|\mathcal{F}_{t-1})\end{array}$$

Backtesting ES

The general protocol for $t \in \mathbb{N}$

- The bank announces ES forecast r_t and VaR forecast z_t
- Decide predictable $\lambda_t(r_t, z_t) \in [0,1]$ ($\Rightarrow$ not shown to the bank)
- Observe realized loss L_t
- Obtain the sequential e-values $X_t = e_p^{\mathrm{ES}}(L_t, r_t, z_t)$
- Compute the e-process ($E_0 = 1$)

$$E_t(\boldsymbol{\lambda}) = (1 - \lambda_t + \lambda_t X_t) E_{t-1}(\boldsymbol{\lambda}) = \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s X_s).$$

Backtesting ES

Theorem 4

Under H_0 *or* H_0', $(E_t(\boldsymbol{\lambda}))_{t=1,\dots,T}$ *is a supermartingale, and*

$$\mathbb{P}\left(\sup_{t\geqslant 1} E_t(\boldsymbol{\lambda}) \geqslant \frac{1}{\alpha}\right) \leqslant \alpha.$$

Our method

- Completely model free
- Anytime validity: one can stop at any stopping time
- Early warning: one can reject at a low threshold such as 2

Comparison with existing methods

Literature	Parametric or dependence assumptions	Forecast structural assumptions	Fixed sample size	Asymptotic test	Reliance on VaR forecast
MF00	yes	yes	yes	yes	yes
AS14	yes	yes	yes	yes	yes
DE17	yes	yes	yes	yes	yes
NZ17	yes	yes	yes	yes	yes
BD22	yes	yes	yes	yes	no
HD22	yes	yes	no	no	yes
This paper	no	no	no	no	yes

Table: Comparison of backtesting methods for ES; parametric or dependence assumptions refer to those on loss distributions, time series models, stationarity, or strong mixing; forecast structural assumptions refer to requirements on the forms and properties of risk forecasts

Backtesting risk measures

- A model-free e-statistic $e : \mathbb{R}^2 \to [0, \infty]$ for (ρ, ϕ) testing ρ
- ρ forecast r_t; ϕ forecast z_t; realized loss L_t; e-value $X_t = e(L_t, r_t, z_t)$

Hypothesis to test

$$H_0 : \begin{array}{c}\text{conditional on } \mathcal{F}_{t-1}: \\ r_t \geqslant \rho(L_t|\mathcal{F}_{t-1}) \text{ and } z_t = \phi(L_t|\mathcal{F}_{t-1})\end{array} \quad \text{for } t = 1, \ldots, T$$

- Decide a (predictable) $\lambda_t(r_t, z_t) \in [0, 1]$
- Compute the test martingale $(E_0 = 1)$

$$E_t(\boldsymbol{\lambda}) = (1 - \lambda_t + \lambda_t X_t) E_{t-1}(\boldsymbol{\lambda}) = \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s X_s).$$

- Size-α test for H_0: reject if $\sup_t E_t(\boldsymbol{\lambda}) \geqslant 1/\alpha$

Choosing λ_t

- Heuristic choice of constant $\lambda_t = \lambda \in [0, 1]$, e.g., $\lambda = 0.01$
- Adaptive choices:
  - Dependent on observed loss data
  - Dependent on forecast
  - Dependent on past forecast
- E-power (Vovk/W.'23) of an e-variable E for an alternative Q:

$$\mathbb{E}^{Q}[\log E]$$

In our setting (where Q_t is unknown):

$$\mathbb{E}^{Q_t}\left[\log(1 - \lambda_t + \lambda_t e(L_t, r_t, z_t)) \mid \mathcal{F}_{t-1}\right]$$

- $\lambda \mapsto \log(1 - \lambda + \lambda e(L_t, r_t, z_t))$ is concave

Choosing λ_t

Fix $\gamma \in (0,1)$; $\gamma = 1/2$ works well

- GRO (growth-rate optimal): $L \sim Q_t$, Grünwald/de Heide/Koolen'23

$$\lambda_t^{\text{GRO}} = \lambda_t^{\text{GRO}}(r,z) = \underset{\lambda\in[0,\gamma]}{\arg\max}\, \mathbb{E}^{Q_t}[\log(1-\lambda+\lambda e(L,r,z)) \mid \mathcal{F}_{t-1}], \quad r,z \in \mathbb{R}^d$$

- GREE (growth-rate for empirical e-statistics):

$$\lambda_t^{\text{GREE}} = \underset{\lambda\in[0,\gamma]}{\arg\max}\, \frac{1}{t-1}\sum_{s=1}^{t-1} \log(1-\lambda+\lambda e(L_s,r_s,z_s))$$

- GREL (growth-rate for empirical losses):

$$\lambda_t^{\text{GREL}} = \lambda_t^{\text{GREL}}(r,z) = \underset{\lambda\in[0,\gamma]}{\arg\max}\, \frac{1}{t-1}\sum_{s=1}^{t-1} \log(1-\lambda+\lambda e(L_s,r,z)), \quad r,z \in \mathbb{R}^d$$

- GREM (GRE mixture): $E_t(\boldsymbol{\lambda}^{\text{GREM}}) = (E_t(\boldsymbol{\lambda}^{\text{GREL}}) + E_t(\boldsymbol{\lambda}^{\text{GREE}}))/2$

Optimality of a method

Definition (Asymptotic optimality)

For $(L_{t-1}, r_t, z_t)_{t\in\mathbb{N}}$ adapted to $(\mathcal{F}_{t-1})_{t\in\mathbb{N}}$ and a given model-free e-statistic e,

- two betting processes $\boldsymbol{\lambda} = (\lambda_t)_{t\in\mathbb{N}}$ and $\boldsymbol{\lambda}' = (\lambda'_t)_{t\in\mathbb{N}}$ are asymptotically equivalent, denoted by $\boldsymbol{\lambda} \simeq \boldsymbol{\lambda}'$, if

$$\frac{1}{T}(\log E_T(\boldsymbol{\lambda}) - \log E_T(\boldsymbol{\lambda}')) \xrightarrow{\text{p}} 0 \quad \text{as } T \to \infty;$$

- a betting process $\boldsymbol{\lambda}$ is asymptotically optimal (AO) if $\boldsymbol{\lambda} \simeq (\lambda_t^{\text{GRO}}(r_t, z_t))_{t\in\mathbb{N}}$.

- The long-term growth rates of the two resulting e-processes are the same
- GRO as the oracle benchmark
- $\psi^*(\mathcal{M}) = \{(r, z)$ in the range of (ρ, ϕ) such that $e(x, r, z) < \infty$ for all $x\}$

Optimality of GREE and GREL

Assumption 1

For all $(r, z) \in \psi^*(\mathcal{M})$, $\sup_{t\in\mathbb{N}} \mathbb{E}^{Q_t}[\log(e(L_t, r, z))] < \infty$.

Theorem 5

Suppose that $(r_t, z_t)_{t\in\mathbb{N}}$ *takes values in* $\psi^*(\mathcal{M})$, $(L_{t-1}, r_t, z_t)_{t\in\mathbb{N}}$ *is adapted to* $(\mathcal{F}_{t-1})_{t\in\mathbb{N}}$, *e is a model-free e-statistic, and Assumption 1 holds.*

(i) $\boldsymbol{\lambda}^{\mathrm{GREE}}$ *is AO if* $(e(L_t, r_t, z_t))_{t\in\mathbb{N}}$ *is iid and* $(r_t, z_t)_{t\in\mathbb{N}}$ *is deterministic.*

(ii) $\boldsymbol{\lambda}^{\mathrm{GREL}}$ *is AO if* $(L_t)_{t\in\mathbb{N}}$ *is iid and either:*

  (a) $(r_t, z_t)_{t\in\mathbb{N}}$ *takes finitely many possible values in* $\mathbb{R}^2$;

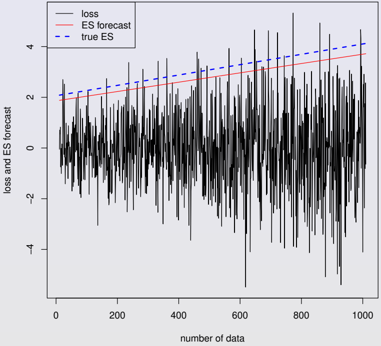
  (b) (r_t, z_t) *lives in a common compact set,* $e(x, r, z)$ *is continuous in* (r, z), *and* $(r_t, z_t) \xrightarrow{\mathrm{p}} (r_0, z_0)$ *as* $t \to \infty$ *for some* (r_0, z_0).

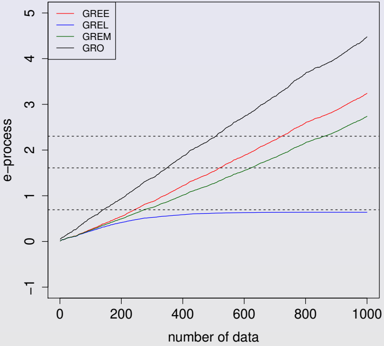
(iii) $\boldsymbol{\lambda}^{\mathrm{GREM}}$ *is AO if either* $\boldsymbol{\lambda}^{\mathrm{GREE}}$ *or* $\boldsymbol{\lambda}^{\mathrm{GREL}}$ *is AO.*

Progress

Example 1: Co-movements of losses and forecasts (linear growth)

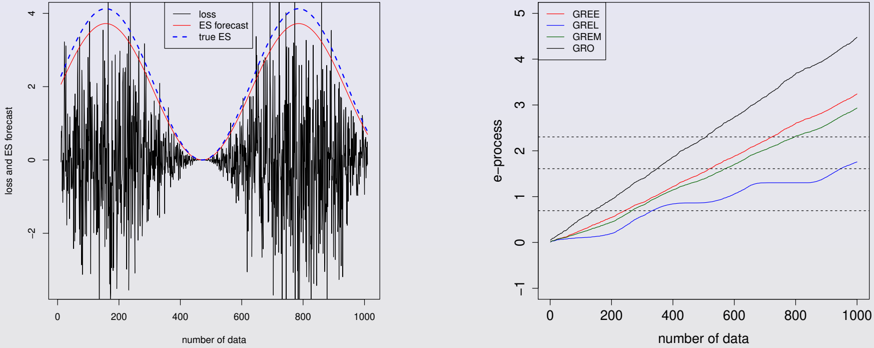
- Sample size for testing $n = 1,000$; size of training data $l = 10$
- Losses: $L_t = Z_t(1 + t/(n + l))$; $\{Z_t\}_{t=1,...,n+l}$ are iid samples from N(0, 1)
- ES forecasts: $r_t = 1.86(1 + t/(n + l))$; VaR forecasts: $z_t = 1.48(1 + t/(n + l))$





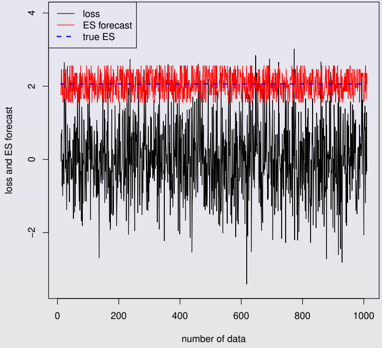
Example 2: Co-movements of losses and forecasts (varying magnitude)

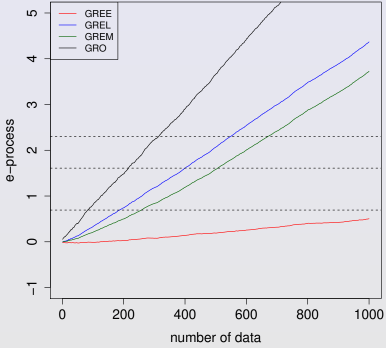
- Losses: $L_t = Z_t(1 + \sin(\theta t))$, $\theta = 0.01$
- ES forecasts: $r_t = 1.86(1 + \sin(\theta t))$; VaR forecasts: $z_t = 1.48(1 + \sin(\theta t))$



Example 3: Forecasts with an estimation error

- Losses: iid $L_t \sim \mathrm{N}(0,1)$
- ES forecasts: $r_t = 2.06 + \varepsilon_t$; VaR forecasts: $z_t = 1.64 + \varepsilon_t$; $\{\varepsilon_t\}_{t=1,\dots,n+l}$ are iid samples uniformly distributed on support $\{\pm i/10 : i = 0, \dots, 5\}$





Time-series model

Data generating process (Nolde/Ziegel'17)

- AR(1)-GARCH(1, 1) process:

$$L_t = \mu_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t Z_t,$$

$$\mu_t = -0.05 + 0.3 L_{t-1}, \quad \sigma_t^2 = 0.01 + 0.1 \varepsilon_{t-1}^2 + 0.85 \sigma_{t-1}^2$$

- The innovations $\{Z_t\}_{t \in \mathbb{N}_+}$ are iid skewed-t with shape parameter $\nu = 5$ and skewness parameter $\gamma = 1.5$
- simulate 1,000 daily losses in each run (1,000 runs)

Time-series model

Forecasters

- Fit AR(1)-GARCH(1, 1) everyday with a moving window of 500 days
- Innovations: normal, t and skewed-t
- Strategies: under-report, point forecast, over-report

Average point forecast over 500 days

	$\widehat{\mathrm{VaR}}_{0.95}$	$\widehat{\mathrm{VaR}}_{0.99}$	$\widehat{\mathrm{VaR}}_{0.875}$	$\widehat{\mathrm{ES}}_{0.875}$	$\widehat{\mathrm{VaR}}_{0.975}$	$\widehat{\mathrm{ES}}_{0.975}$
normal	0.605	0.883	0.403	0.606	0.734	0.888
t	0.528	0.974	0.300	0.566	0.709	1.034
skewed-t	0.658	1.217	0.365	0.701	0.888	1.281
true	0.658	1.242	0.359	0.706	0.897	1.312

Backtesting ES (e-process), GREM

E-value rejection thresholds: 2, 5, and 10

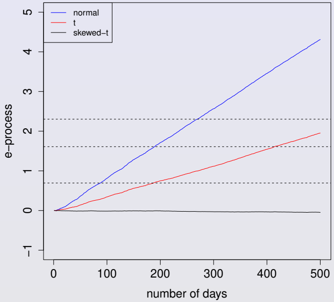


Figure: Average (Log) e-processes testing $ES_{0.975}$ with respect to number of days using the GREM method

Backtesting ES, GREM

	−10% ES			exact			+10% ES		
Threshold	2	5	10	2	5	10	2	5	10
normal	99.8	99.5	98.5	99.3	95.7	88.3	94.8	79.8	62.1
t	98.4	88.8	77.1	88.1	63.9	43.1	70.0	34.9	15.6
skewed-t	47.6	16.1	6.2	18.8	4.0	0.8	7.9	1.1	0.1

Table: Percentage of rejections (%) for $\mathrm{ES}_{0.975}$ forecasts using the GREM method within the total $1,000$ trials

Detection of structural change

Data generating process (Hoga/Demetrescu'22)

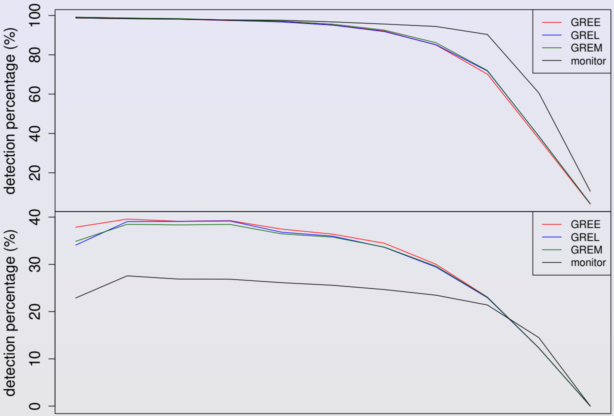
- GARCH(1, 1) process:

$$L_t = -\sigma_t Z_t, \quad \sigma_t^2 = 0.00001 + 0.04\varepsilon_{t-1}^2 + \beta_t \sigma_{t-1}^2$$

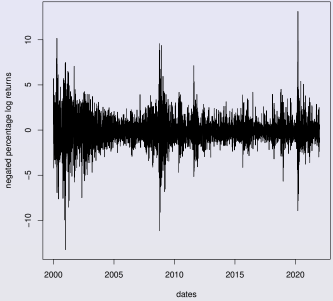
- The innovations $\{Z_t\}_{t\in\mathbb{N}_+}$ are iid skewed-t with shape parameter $\nu = 5$ and skewness parameter $\gamma = 0.95$
- Simulate 250 daily losses for forecasting and 250 for testing
- $\beta_t = 0.7 + 0.25\mathbb{1}_{\{t>b*\}}$; $b^* + 1$ is the time where structural change happens

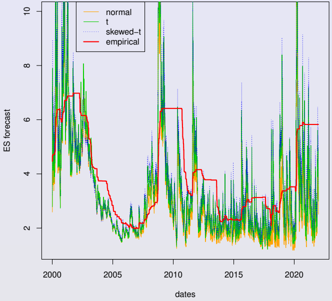
Backtesting $\text{ES}_{0.95}$

Percentage of detections (%) and average number of days needed to detect structural change (ARL) with respect to b^*; black line ("monitor") represents the result of the sequential monitoring method in Hoga/Demetrescu'22



Data analysis setting



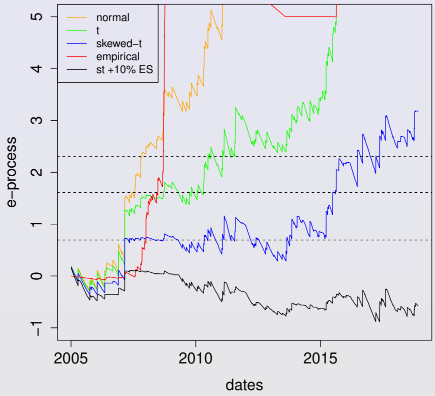


- Negated log-returns of the NASDAQ Composite index from Jan 16, 1996 to Dec 31, 2021
- Fitted to an AR(1)-GARCH(1, 1) model with moving window of 500
- Sample size after initial training: $n = 5,536$

Jan 2005 - Dec 2021, GREM, $ES_{0.975}$

Threshold	2	5	10	
normal	540	610	713	(30.28)
t	540	933	1381	(10.25)
skewed-t	540	2639	2889	(4.169)
st +10% ES	–	–	–	(−0.6896)
empirical	756	862	931	(8.454)

Table: Number of days taken to reject the $ES_{0.975}$ forecasts, and final log e-values (in brackets); "–" means no rejection



Future directions

- E-backtesting other risk measures
  - Gini deviation ($\frac{1}{2}\mathbb{E}[|X - X'|]$): Model-free e-statistics take the form of $r \mapsto \frac{|x_1 - x_2|}{2r}$ (requires two iid copies)
  - Distortion risk measures
- Game theoretic framework
  - Financial institution: report as low risk forecasts as possible
  - Regulator: reject when e-process becomes large
  - Equilibrium risk forecasts and betting process $(\lambda_t)_{t\in\mathbb{N}}$
- Other methods choosing betting process $(\lambda_t)_{t\in\mathbb{N}}$
  - Optimal betting process for specific distributions
  - Optimal betting process for general dependence structures

Thank you

Thank you for your attention

https://arxiv.org/abs/2209.00991