Merging e-values via martingales and e-backtesting

Ruodu Wang

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

Department of Statistics and Actuarial Science
University of Waterloo

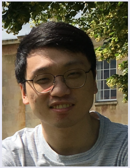
Workshop: "Safe, Anytime-Valid Inference (SAVI) and Game-theoretic Statistics"
Eindhoven, Netherlands, May 31, 2022

Agenda

1. E-values
2. Merging sequential e-values
3. Merging independent e-values
4. Merging dependent e-values and the e-BH procedure
5. Risk forecasts and backtests

E-values




Aaditya Ramdas (Carnegie Mellon)	Vladimir Vovk (Royal Holloway)	Bin Wang (CAS Beijing)	Qiuqi Wang (Waterloo)	Johanna F. Ziegel (Bern)

- Vovk/W., E-values: Calibration, combination, and applications. Annals of Statistics, 2021, arXiv:1912.06116
- Vovk-Wang-W., Admissible ways of merging p-values under arbitrary dependence. Annals of Statistics, 2022, arXiv:2007.14208
- Vovk/W., Merging sequential e-values via martingales. 2022, arXiv:2007.06382
- W./Ramdas, False discovery rate control with e-values. JRSSB, 2022, arXiv:2009.02824
- Wang/W./Ziegel, E-statistics, model-free tests, and backtesting the Expected Shortfall. 2022, working paper

What is an e-value?

- A hypothesis H: a set of probability measures

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

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

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

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

- For simple hypothesis $\{\mathbb{P}\}$
  - precise e-variable: random variable ≥ 0 with mean 1
  - precise e-process: supermartingale ≥ 0 with initial value 1

What is an e-value?

- A p-variable for testing H is a random variable $P : \Omega \to [0, \infty)$ that satisfies $\sup_{Q \in H} Q(P \le \alpha) \le \alpha$ for all $\alpha \in (0, 1)$
- 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
- An e-process has retrospective validity (Ville's inequality):

$$\mathbb{P}\left(\sup_{t \ge 0} X_t \ge \frac{1}{\alpha}\right) \le \alpha \implies \inf_{t \ge 0} X_t^{-1} \text{ is a p-value}$$

- Bayes factors (simple hypothesis) and likelihood ratios:

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

E for Expectation (or Evidence)

	requirement	specific interpretation	representative forms	keyword
p-value P	$\mathbb{P}(P \le \alpha) \le \alpha$ for $\alpha \in (0,1)$	probability of a more extreme observation	$\mathbb{P}(T' \le T(\mathbf{X}) \mid \mathbf{X})$	(conditional) probability
e-value E	$\mathbb{E}^{\mathbb{P}}[E] \le 1$ and $E \ge 0$	likelihood ratios, stopped martingales, and betting scores	$\mathbb{E}^{\mathbb{P}}\left[\frac{\mathrm{d}\mathbb{Q}}{\mathrm{d}\mathbb{P}} \middle\vert \mathbf{X}\right]$ $\mathbb{E}^{\mathbb{P}}[M_\tau \mid \mathbf{X}]$	(conditional) expectation

An analogy of p-variables and e-variables for a simple hypothesis $\{\mathbb{P}\}$

- $\mathbf{X}$ is data
- $T(\mathbf{X})$ is any test statistic
- T' is an independent copy of $T(\mathbf{X})$ under $\mathbb{P}$
- $\mathbb{Q}$ is any probability measure
- M is a test supermartingale under $\mathbb{P}$ and τ a stopping time

(not to be confused with other objects bearing the name of e-values)

Example in testing multiple hypotheses

Multi-armed bandit problems

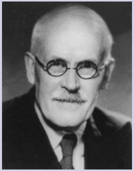
- K arms
- null hypothesis k: arm k has mean reward at most 1
- strategy (k_t): at time $t \geq 1$, pull arm k_t, obtain an iid reward $X_{k_t,t} \geq 0$
- aim: quickly detect arms with mean > 1
  - or maximize profit, minimize regret, etc ...
- running reward: $M_{k,t} = \prod_{j=1}^{t} X_{k,j} \mathbb{1}_{\{k_j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

Calibration

- Admissible p-to-e calibrators
  - Power calibrators: $f_\kappa(p) = \kappa p^{\kappa-1}$ for $\kappa \in (0,1)$
  - Shafer's: $f(p) = p^{-1/2} - 1$
  - Averaging f_κ: $\int_0^1 \kappa p^{\kappa-1} \mathrm{d}\kappa = \frac{1-p+p\ln p}{p(-\ln p)^2}$
- the only admissible e-to-p calibrator: $e \to e^{-1} \wedge 1$

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.)



Sequential e-values

E-variables $E_1, \ldots, E_K$ are sequential if E_k is an e-variable conditional on $E_1, \ldots, E_{k-1}$ for each k.

- $\mathbb{E}[E_k \mid E_1, \ldots, E_{k-1}] \le 1$ for all $k \in [K] := \{1, \ldots, K\}$
- E-values $e_1, \ldots, e_K$ are obtained by laboratories $1, \ldots, K$
- Laboratory k makes sure that its result e_k is a valid e-value given the previous results $e_1, \ldots, e_{k-1}$
- Independent e-variables are sequential

Progress

One little philosophical slide

P-values can be avoided if one only aims for a binary decision

- If $\alpha = 0.05$ is set a priori, then $p = 0.049$ and $p = 0.001$ carry the same significance
- If α is not set a priori, then we cannot reject anything after we see the data
- When we need to operate on p-values, the abstract p-value becomes convenient
    - p-combination, Bonferroni, closed testing, FDR (Benjamini-Hochberg), FCR, meta analysis ...

One little philosophical slide

P-values can be avoided if one only aims for a binary decision

- If $\alpha = 0.05$ is set a priori, then $p = 0.049$ and $p = 0.001$ carry the same significance
- If α is not set a priori, then we cannot reject anything after we see the data
- When we need to operate on p-values, the abstract p-value becomes convenient
  - p-combination, Bonferroni, closed testing, FDR (Benjamini-Hochberg), FCR, meta analysis ...

Same for e-values?

- The abstract notion is needed when we operate on e-values
  - e-combination, e/p-calibration, closed testing, FDR (e-BH), FCR (e-BY), meta analysis ...
  - out-come level, study level, or multiple hypotheses
- We do not specify how they are obtained or the target statistical problem

Handling e-values

- We are supplied with K e-values for a hypothesis H_0
  - Obtained from other papers/talks ...
  - They may be sequential, independent, or arbitrarily dependent

How do we come up with one output e-value?

Handling e-values

- We are supplied with K e-values for a hypothesis H_0
  - Obtained from other papers/talks ...
  - They may be sequential, independent, or arbitrarily dependent

How do we come up with one output e-value?

Definition (e/ie/se-merging functions)

An e-merging/ie-merging/se-merging function is a Borel function $F : [0, \infty)^K \to [0, \infty)$ such that $F(E_1, \dots, E_K)$ is an e-variable for all/all independent/all sequential e-variables $E_1, \dots, E_K$.

$$\{\text{e-merging}\} \subsetneq \{\text{se-merging}\} \subsetneq \{\text{ie-merging}\}$$

Sequential vs independent e-variables

- An iid sample $(X_1, \dots, X_K)$ from $\theta_{\rm tr} \in \Theta$ are sequentially revealed
- Test $H_0 : \theta_{\rm tr} = 0$ against $H_1 : \theta_{\rm tr} \in \Theta_1$ where $0 \notin \Theta_1 \subseteq \Theta$.
  - It does not hurt to think about testing $\mathrm{N}(\theta, 1)$
- Let ℓ be the likelihood ratio function

$$\ell(x; \theta) = \frac{\mathrm{d}Q_\theta}{\mathrm{d}Q_0}(x),$$

  where Q_θ is the probability measure corresponds to $\theta \in \Theta$
- $\ell(X_k; \theta)$ for any $\theta \in \Theta$ and $k \in [K]$ is an e-variable for H_0

Sequential vs independent e-variables

The scientist may choose two difference strategies:

(a) Fix $\theta_1, \dots, \theta_K \in \Theta_1$

- One may simply choose all θ_k to be the same

(b) Adaptively update $\theta_1, \dots, \theta_K$, where θ_k is estimated from $(X_1, \dots, X_{k-1})$ for each k.

- E.g., Bayesian update or point estimates

In either case:

- Define the e-variables $E_k := \ell(X_k; \theta_k)$ for $k \in [K]$
  - In (a), $E_1, \dots, E_K$ are independent e-variables
  - in (b), $E_1, \dots, E_K$ are sequential e-variables
- Combine $(E_k)_{k \in [K]}$ to get an output e-variable, e.g., $\prod_{k=1}^{K} E_k$

Sequential vs independent e-variables

- An iid sample $(X_1, \dots, X_K)$ from $\mathrm{N}(\theta_{\mathrm{tr}}, 1)$
- $H_0 : \theta_{\mathrm{tr}} = 0$ against $H_1 : \theta_{\mathrm{tr}} > 0$
- Set $\theta_{\mathrm{tr}} = 0.3$
- Five ways to obtain $E_k = \ell(X_k; \theta_k)$
  - (i) $\theta_k = \theta_{\mathrm{tr}} = 0.3$: true alternative, growth-optimal
  - (ii) $\theta_k = \theta_0 = 0.1$: misspecified alternative
  - (iii) θ_k follows an iid uniform distribution on $[0, 0.5]$
  - (iv) θ_k follows a Bayesian update rule with a prior $\theta \sim \mathrm{N}(\theta_0, 0.2^2)$
  - (v) θ_k is MLE based on $(X_1, \dots, X_{k-1})$ with $\theta_1 = \theta_0$
- (i)-(iii): independent e-variables; (iv)-(v): sequential
- Report $\prod_{k=1}^K E_k$

Sequential vs independent e-variables

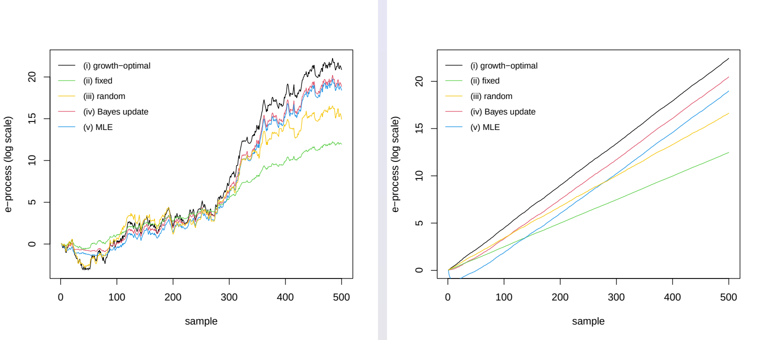


Figure: A few ways of constructing e-processes from likelihood ratio. Left: one run; Right: the average (log) of 1000 runs.

- Trade-off: sequential vs independent

Merging with U-statistics

The U-statistics for $n \in \{0, 1, \dots, K\}$:

$$U_n(e_1, \dots, e_K) := \frac{1}{\binom{K}{n}} \sum_{\{k_1,\dots,k_n\} \subseteq \{1,\dots,K\}} e_{k_1} \dots e_{k_n}.$$

- product ($n = K$)
- arithmetic average M_K ($n = 1$)
- constant 1 ($n = 0$)

Proposition 1

Each of the U-statistics and their convex mixtures is an admissible ie-merging function and an admissible se-merging function.

- Admissibility: not strictly dominated by any

Merging with U-statistics

Proposition 2

For the product function $P_K : (e_1, \ldots, e_K) \mapsto \prod_{k=1}^K e_k$ *and any ie-merging function* F*, it holds*

$$(e_1, \ldots, e_K) \in [1, \infty)^K \implies F(e_1, \ldots, e_K) \le P_K(e_1, \ldots, e_K).$$

Merging with U-statistics

Proposition 2

For the product function $P_K : (e_1, \dots, e_K) \mapsto \prod_{k=1}^K e_k$ *and any ie-merging function* F*, it holds*

$$(e_1, \dots, e_K) \in [1, \infty)^K \implies F(e_1, \dots, e_K) \le P_K(e_1, \dots, e_K).$$

In the setting that all e-variables are independent and have mean ≥ 1 under the alternative, the product function P_K is

- uniformly "the most powerful" among all ie-merging functions
  - largest expected value under the alternative
- uniformly "the least stable" among all se-merging functions
  - largest second moment under the alternative

Merging independent e-values

Example

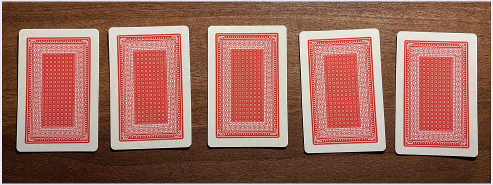
The function

$$(e_1, e_2) \mapsto \frac{1}{2}\left(\frac{e_1}{1+e_1} + \frac{e_2}{1+e_2}\right)(1 + e_1 e_2)$$

is

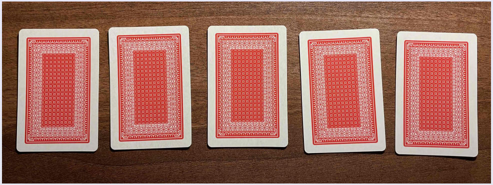
- an admissible ie-merging function;
- not a convex mixture of U-statistics;
- not an se-merging function.

Sequential e-merging

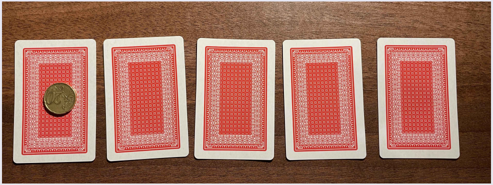


(each card has an e-value face down)

Sequential e-merging

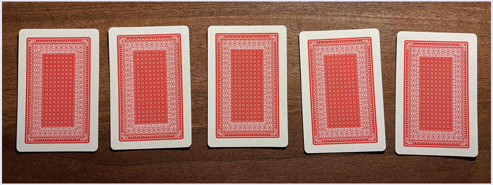


(each card has an e-value face down)

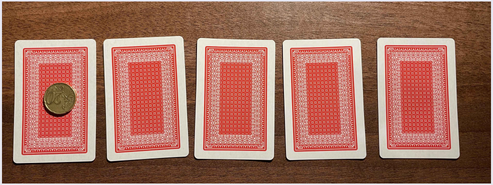


(bet on the first card)

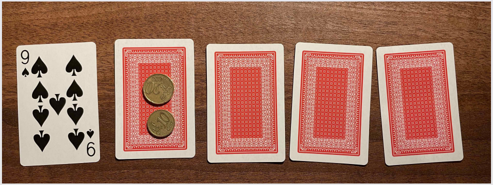
Sequential e-merging



(each card has an e-value face down)



(bet on the first card)



(reveal the card and proceed)

Sequential e-merging

- $\mathbf{e} := (e_1, \ldots, e_K)$; $\mathbf{e}_{(k)} := (e_1, \ldots, e_k)$; $\mathbf{e}_{(0)} := \varnothing$
- For some functions $\lambda_1, \ldots, \lambda_K$, define $S_0 = 1$ and

$$S_k(\mathbf{e}) = \prod_{i=1}^{k} \left(1 - \lambda_j(\mathbf{e}_{(j-1)})(e_j - 1)\right), \quad k \in [K]$$

  The sequence of functions $(S_k)_{k\in\{0,1,\ldots,K\}}$ is a test martingale
- $(S_k(\mathbf{E}))_{k\in\{0,1,\ldots,K\}}$ is an e-process
- Define the martingale merging function $F(\mathbf{e}) = S_K(\mathbf{e})$
- F and S_k are connected via

$$S_k(e_1, \ldots, e_K) = F(e_1, \ldots, e_k, 1, \ldots, 1).$$

- F is generally not monotone

Sequential e-merging

Theorem 1

(i) *A convex combination of martingale e-merging functions is a martingale e-merging function.*

(ii) *A martingale e-merging function is an se-merging function.*

(iii) *Each se-merging function is dominated by a martingale e-merging function.*

- Arithmetic average, product, and U-statistics are all special cases of martingale e-merging functions

Sequential e-merging

- An e-variable E is precise if $\mathbb{E}[E] = 1$

Theorem 2

For a sequence of functions $F = (F_k)_{k=1,\dots,K}$, equivalent are:

(i) *F is a test martingale;*

(ii) *$F(\mathbf{E})$ is a martingale (wrt. the natural filtration of $\mathbf{E}$) for any vector $\mathbf{E}$ of precise and sequential e-values;*

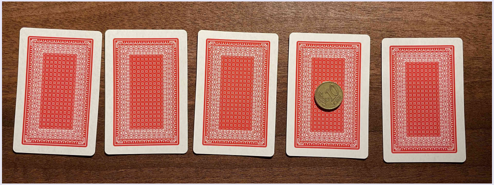
(iii) *F is anytime valid and precise; i.e., it satisfies*

  (a) *$F_\tau(\mathbf{E})$ is an e-variable for any vector $\mathbf{E}$ of sequential e-values and any stopping time τ;*

  (b) *For each $k \in [K]$, $\mathbb{E}[F_k(\mathbf{E})] = 1$ for any vector $\mathbf{E}$ of precise and sequential e-variables.*

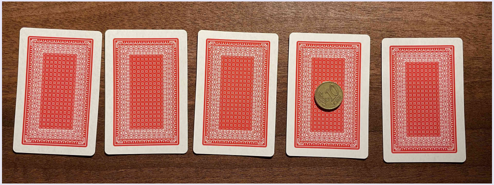
Progress

Independent e-merging

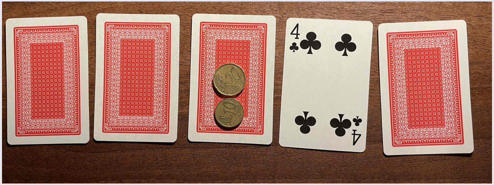


(choose a card to bet on)

Independent e-merging

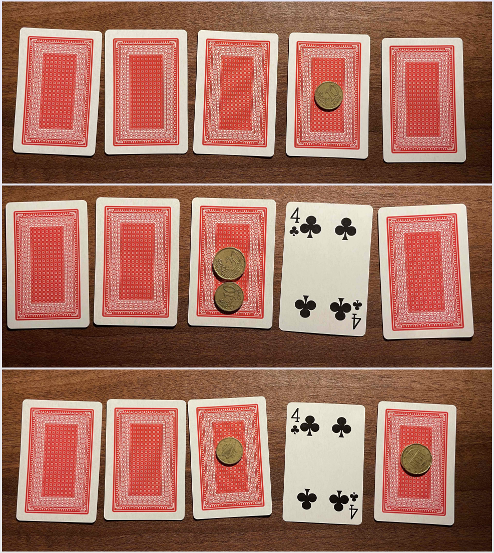


(choose a card to bet on)



(choose both the next card and the bet)

Independent e-merging



(choose a card to bet on)

(choose both the next card and the bet)

(one could also bet several cards simultaneously - mixed strategy)

Independent e-merging

- Write $\mathbf{e}^{\pi}_{(k)} = (e_{\pi_1}, \dots, e_{\pi_k})$ where π_j may be a function
- A reading strategy $\pi = (\pi_k)_{k \in [K]}$ is such that
  - $\pi_k : [0, \infty)^{k-1} \to [K]$
  - $\pi_k(\mathbf{e}^{\pi}_{(k-1)}) \neq \pi_j(\mathbf{e}^{\pi}_{(j-1)})$ for all $\mathbf{e} \in [0, \infty)^K$ and $j \neq k$; i.e., you can only read the same e-value once

Lemma 1

Let $E_1, \dots, E_K$ be independent e-variables, and π a reading strategy. Recursively define $E^{\pi}_k = E_{\pi_k(E^{\pi}_1, \dots, E^{\pi}_{k-1})}$ for $k \in [K]$. Then $E^{\pi}_1, \dots, E^{\pi}_K$ are sequential e-variables. If $E_1, \dots, E_K$ are iid, then so are $E^{\pi}_1, \dots, E^{\pi}_K$.

Independent e-merging

- A reordered test martingale: $S_0 = 1$,

$$S_k^{\lambda,\pi}(\mathbf{e}) = \prod_{i=1}^{k} \left(1 + \lambda_j(\mathbf{e}^{\pi}_{(j-1)})(e_{\pi_j(\mathbf{e}_{(j-1)})} - 1)\right), \quad k \in [K]$$

- A generalized martingale merging function (GMMF) is a mixture of $S_K^{\lambda,\pi}$ above

Proposition 3

Any GMMF is an ie-merging function.

Independent e-merging

- A reordered test martingale: $S_0 = 1$,

$$S_k^{\lambda,\pi}(\mathbf{e}) = \prod_{i=1}^{k} \left(1 + \lambda_j(\mathbf{e}^{\pi}_{(j-1)})(e_{\pi_j(\mathbf{e}_{(j-1)})} - 1)\right), \quad k \in [K]$$

- A generalized martingale merging function (GMMF) is a mixture of $S_K^{\lambda,\pi}$ above

Proposition 3

Any GMMF is an ie-merging function.

- Are all ie-merging function dominated by some GMMF, like se-merging functions dominated by MMF?

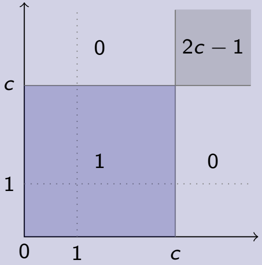
Merging independent e-values

Example

Fix a constant $c > 1$ and define the function $G : [0,\infty)^2 \to \mathbb{R}$ by

$$G(\mathbf{e}) = \mathbb{1}_{[0,c)^2}(\mathbf{e}) + (2c-1)\mathbb{1}_{[c,\infty)^2}(\mathbf{e}).$$

- G is an ie-merging function
- G is not dominated by any GMMF
- G is not increasing or precise



This counter-example is provided by Zhenyuan Zhang

Merging independent e-values

Open questions:

- What are all (precise, increasing) ie-merging functions?
- Does the set of precise ie-merging functions coincide with GMMF?
- Are there useful ie-merging functions beyond GMMF?
- After all, what is the value of independence (if any)?

Progress

Arbitrarily dependent e-values

Theorem 3

Suppose that F is a symmetric e-merging function. Then $F \le \lambda + (1-\lambda)M_K$ for some $\lambda \in [0,1]$, and F is admissible if and only if $F = \lambda + (1-\lambda)M_K$ with $\lambda = F(\mathbf{0})$.

- For any symmetric e-merging function F:

$$F(\mathbf{e}) > 1 \implies M_K(\mathbf{e}) \ge F(\mathbf{e}).$$

- Asymmetric e-merging: $\mathbf{e} \mapsto \boldsymbol{\lambda} \cdot \mathbf{e}$ for $\boldsymbol{\lambda} \in \Delta_K$ where Δ_K is the standard K-simplex

---

Vovk-W., E-values: Calibration, combination, and applications. Annals of Statistics, 2021, Theorem 3.2 (relaxing monotonicity: Proposition E.3)

Connection to p-merging

Theorem 4

For any admissible p-merging function F and $\epsilon \in (0,1)$, there exist $(w_1, \ldots, w_K) \in \Delta_K$ and admissible calibrators $f_1, \ldots, f_K$ such that

$$F(\mathbf{p}) \le \epsilon \iff \sum_{k=1}^{K} w_k f_k(p_k) \ge \frac{1}{\epsilon}.$$

If F is symmetric, then there exists an admissible calibrator f such that

$$F(\mathbf{p}) \le \epsilon \iff \frac{1}{K} \sum_{k=1}^{K} f(p_k) \ge \frac{1}{\epsilon}.$$

Vovk-Wang-W., Admissible ways of merging p-values under arbitrary dependence. Annals of Statistics, 2022, Theorem 5.1

E-BH procedure

- $e_1, \dots, e_K$: e-values associated to $H_1, \dots, H_K$, respectively
- $e_{[1]} \geq \cdots \geq e_{[K]}$: order statistics
- The rough relation $e \sim 1/p \Rightarrow$ use $1/e$ in the BH procedure

E-BH procedure

The e-BH procedure $\mathcal{G}_\alpha : [0,\infty]^K \to 2^{\mathcal{K}}$ for $\alpha > 0$ rejects hypotheses with the largest k^* e-values, where

$$k^* = \max\left\{k \in \mathcal{K} : \frac{ke_{[k]}}{K} \geq \frac{1}{\alpha}\right\}.$$

E-BH procedure

Theorem 5

The e-BH procedure $\mathcal{G}_\alpha$ applied to arbitrary e-values has FDR at most $K_0\alpha/K$.

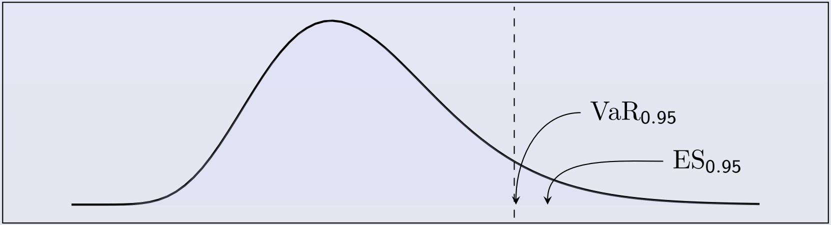
	nice cases	general (AD)
p-BH	$\frac{K_0}{K}\alpha$	penalty
e-BH	boosting	$\frac{K_0}{K}\alpha$

- The catch: for the same data set, $e \leq 1/p$ and often $e < 1/p$

W.-Ramdas, False discovery rate control with e-values. JRSSB, 2022, Theorem 2

Progress

VaR and ES



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

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

$$\mathrm{VaR}_p(X) = q_p(X)$$
$$= \inf\{x \in \mathbb{R} : \mathbb{P}(X \le x) \ge p\}$$

(left-quantile)

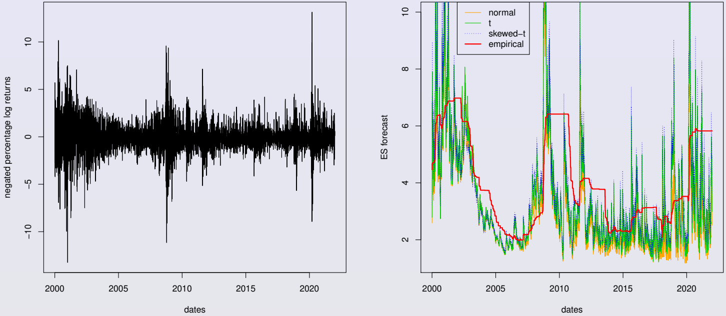
Expected Shortfall (ES), $p \in (0,1)$

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

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

(also: TVaR/CVaR/AVaR)

An example



- Negated log-returns (in %) of the NASDAQ Composite index from Jan, 2000 to Dec 2021
- Fitted (AR(1)-GARCH(1, 1)) or empirical $\mathrm{ES}_{0.975}$ forecasts with moving window of 500

Backtesting risk measures

- Risk measure ρ to backtest
- Define

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

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

Hypothesis to test

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

Some summary

- ES is the standard risk measure in banking
- VaR is easy to backtest and model-free methods are available
- ES is difficult to backtest and no model-free methods are available

E-backtesting ES

Daily observations

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

Hypothesis to test

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

A weaker hypothesis

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

Obtaining sequential e-values

Define the function

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

Theorem 6

For H_0 or H_0', $e_p(L_t, r_t, z_t)$, $t = 1, \dots, T$ are sequential e-variables.

- Proof: based on Rockafellar/Uryasev'02
- e_p is the only choice in this procedure in some sense

Backtesting ES

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

- The bank announces ES forecast r_t and VaR forecast z_t
- Decide predictable $\lambda_t(r_t, z_t) \in [0,1]$
    - Choosing λ_t: many papers/talks ...
- Observe realized loss L_t
- Obtain the e-value $x_t = e_p(L_t, r_t, z_t)$
- Compute the e-process ($E_0 = 1$)

$$E_t = E_{t-1}(1 - \lambda_t + \lambda_t x_t) = \prod_{s=1}^{t} (1 - \lambda_s + \lambda_s x_s).$$

- model free, anytime valid, and allowing for intermediate assessments

Simulation studies

Data generating process (Nolde/Ziegel'17)

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

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

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

- The innovations $\{Z_t\}_{t\in\mathbb{N}_+}$ are iid skew-t with shape parameter $\nu = 5$ and skewness parameter $\gamma = 1.5$
- simulate 5500 daily losses (one run)

Simulation studies

Forecasters

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

Average point forecast over 5000 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)

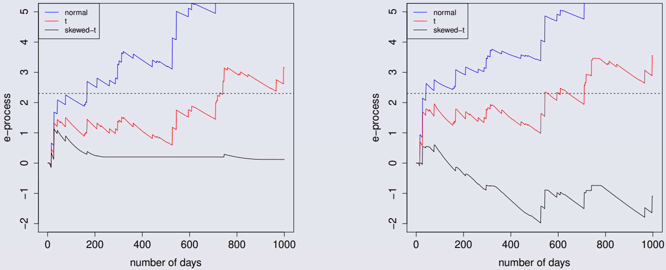


Figure: (Log) e-processes testing $\mathrm{ES}_{0.975}$ with respect to number of days. Left: constant Kelly; right: functional Kelly

Backtesting ES (constant Kelly)

	constant Kelly				
	−10% ES	−10% both	exact	+10% both	+10% ES
normal	42 (58.25)	76 (59.94)	167 (39.70)	313 (23.81)	313 (25.41)
t	296 (33.71)	296 (37.97)	728 (19.03)	1958 (6.417)	1832 (8.665)
skewed-t	1914 (5.490)	1921 (5.497)	– (−0.3122)	– (0.1477)	– (0.06787)

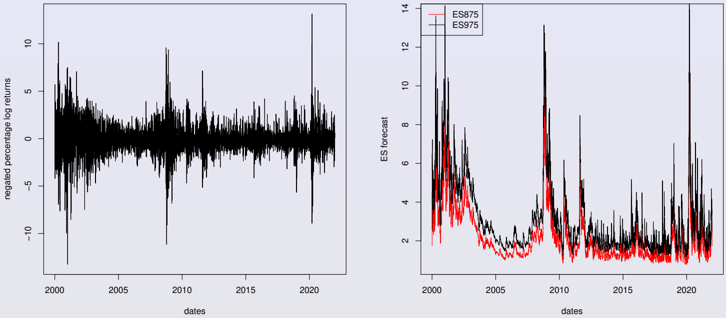
Table: Number of days taken to reject $\mathrm{ES}_{0.975}$ forecasts; "–" means no rejection is detected till day 5000; numbers in brackets are final (log) e-values

Backtesting ES (functional Kelly)

	functional Kelly				
	−10% ES	−10% both	exact	+10% both	+10% ES
normal	27 (50.66)	41 (52.92)	41 (36.93)	42 (24.45)	209 (25.84)
t	167 (31.67)	167 (35.38)	544 (20.32)	1405 (9.477)	1326 (11.71)
skewed-t	1914 (6.370)	1866 (7.185)	– (−1.524)	– (−5.566)	– (−6.044)

Table: Number of days taken to reject $\mathrm{ES}_{0.975}$ forecasts; "–" means no rejection is detected till day 5000; numbers in brackets are final (log) e-values

Empirical setting

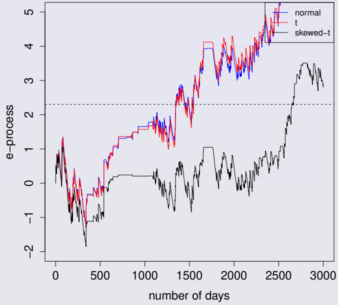


- Negated log-returns of the NASDAQ Composite index from Jan 2000 to Dec 2021
- Fitted to an AR(1)-GARCH(1, 1) model with moving window of 500
- Sample size after initial training: $n = 5,536$

Jan 2005 - Dec 2021, functional Kelly, $\mathrm{ES}_{0.875}$ (log scale)

Impact of financial crisis

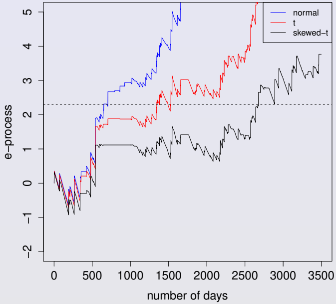
	normal	t	skewed-t
average $\widehat{\mathrm{ES}}_{0.875}$	1.823	1.829	1.965
rejection day	1344	1345	2645
final (log) e-value	14.70	14.91	4.722



Jan 2005 - Dec 2021, functional Kelly, $\mathrm{ES}_{0.975}$ (log scale)

Impact of financial crisis

	normal	t	skewed-t
average $\widehat{\mathrm{ES}}_{0.975}$	2.624	2.979	3.218
rejection day	650	1344	2676
final (log) e-value	23.84	10.56	4.825



Thank you

Working paper series on e-values www.alrw.net/e