Session: "E-Values vs. P-Values: Contrasts and Synergies"

December 16, 2024 @ ICSDS 2024 (Nice, France)

Mind the Filtration: E-Processes vs. P-Processes @ Stopping Times

Yo Joong "YJ" Choe

Data Science Institute, University of Chicago

yjchoe.github.io

Main Reference for This Talk

Yo Joong Choe & Aaditya Ramdas (2024).
"Combining Evidence Across Filtrations."
Preprint: https://arxiv.org/abs/2402.09698


*A Primer on E-Values/E-Processes

E-Value: "E is the New P"

- Given n data points $X_1, \ldots, X_n$ (with a fixed sample size), an **e-value** $E = E_n(X_1, \ldots, X_n)$ for a composite null hypothesis H_0 is a nonnegative random variable satisfying

$$\boxed{\mathbb{E}_{H_0}[E] \leq 1.}$$

- **E-values can be used for testing:** for any $\alpha \in (0, 1)$, by Markov's inequality,

$$P(E \geq 1/\alpha) \leq \alpha, \quad \forall P \in H_0.$$

- **E-values can be combined easily (under arbitrary dependence):** If we have K arbitrarily dependent e-values $E^{(1)}, \ldots, E^{(K)}$ for H_0, their **mean** is also an e-value for H_0:

$$\mathbb{E}_{H_0}\left[\frac{1}{K}\sum_{k=1}^{K} E^{(k)}\right] = \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{H_0}\left[E^{(k)}\right] \leq 1.$$

A key benefit of using e-values over p-values!

cf. Ramdas, Grünwald, Shafer, & Vovk (*Stat. Sci.*, 2023)

Evidence Measures for Sequential Anytime-Valid Inference

- Let $\mathbb{F} = (\mathscr{F}_t)_{t \geq 0}$ be a **filtration**, say, $\mathscr{F}_t = \sigma(X_1, \ldots, X_t)$ (<u>sequentially</u> observed data).

- **Anytime-validity** refers to validity at <u>any</u> **arbitrary (possibly infinite) $\mathbb{F}$-stopping time τ:**

$\mathfrak{e}_t = C(\mathfrak{p}_t)$ **(P-to-E Calibration)**

E-Process $(\mathfrak{e}_t)_{t \geq 0}$

Nonnegative $\mathbb{F}$-process for H_0

$$\mathbb{E}_P[\mathfrak{e}_\tau] \leq 1,$$

for any $P \in H_0$.

P-Process $(\mathfrak{p}_t)_{t \geq 0}$

$[0, 1]$-valued $\mathbb{F}$-process for H_0

$$P(\mathfrak{p}_\tau \leq \alpha) \leq \alpha,$$

for any $P \in H_0$ and $\alpha \in (0, 1)$.

$\mathfrak{p}_t = 1/\sup_{i \leq t} \mathfrak{e}_i$ **(Ville's Inequality)**

*Ramdas et al. (2020)

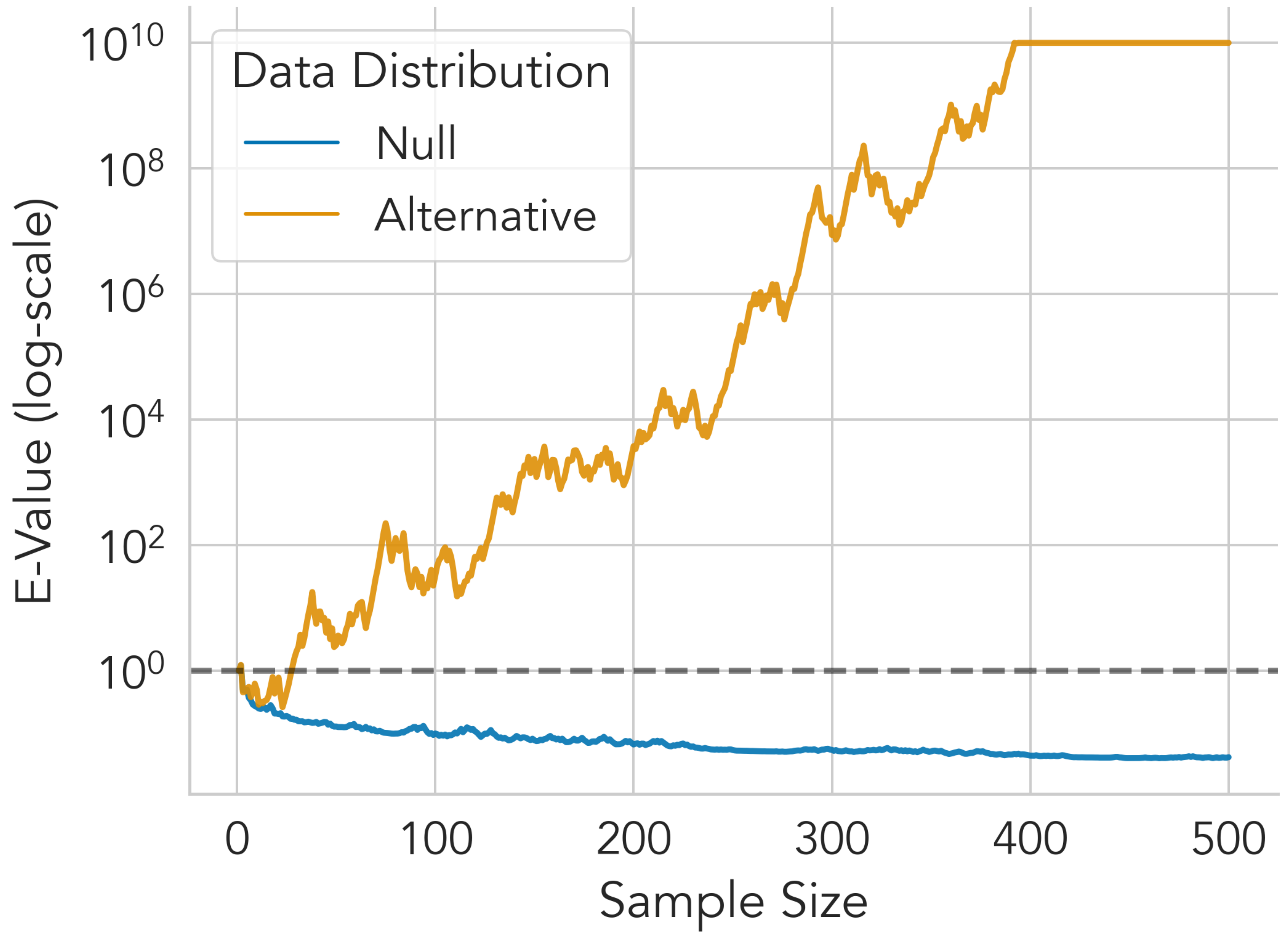
E-Process = Anytime-Valid Evidence Against the Null

E-Process $(e_t)_{t\geq 0}$

Nonnegative $\mathbb{F}$-process for H_0

$$\mathbb{E}_P[e_\tau] \leq 1,$$

for any $P \in H_0$.



An e-process is expected to be small under the ***null****;*
we want it to grow large under the ***alternative****.*

Can We Combine Arbitrary E-*Processes*?

01110011100100010100100001110101100110101001…

Can we test if this sequence is random (i.i.d.) at arbitrary **data-dependent stopping times**?

(e.g., first time we observe five consecutive zeros)

Example: Sequentially Testing Randomness

"Is your data stream actually random?"

- We want to sequentially test whether a binary stream of data $X_1, X_2, \ldots$ is random:

$$H_0^{\mathrm{iid}} : X_1, X_2, \ldots \text{ is i.i.d.}$$

- H_0^{iid} is a family of distributions over the entire sequence: $H_0^{\mathrm{iid}} = \{\mathrm{Ber}(p)^\infty : p \in [0,1]\}$.
- Essentially "equivalent" to testing **exchangeability**. (Ramdas et al., 2022)
- General takeaway translates to non-binary streams as well.

Two Different E-Processes Exist. Can We Combine Them?

	Universal Inference E-Process $(e_t^{UI})_{t \geq 0}$ (Ramdas et al., *IJAR* 2022)	Conformal Test Martingale $(e_t^{conf})_{t \geq 0}$ (Vovk, *Stat. Sci.* 2021)
FORM	$e_t^{UI} = \dfrac{\text{mixture over Markov alternatives}}{\text{maximum likelihood under iid null}}$	$e_t^{conf} = \prod_{i=1}^{t} \left[1 + \lambda \left(p_i - \frac{1}{2}\right)\right], \lambda \in \mathbb{R}$
POWERFUL AGAINST…	**Markov** alternatives (powerless against changepoints)	**Changepoint** alternatives (powerless against Markov)
FILTRATION & STOPPING	The data filtration $\mathscr{F}_t = \sigma(X_1, \ldots, X_t)$ **(allows data-dependent stopping)**	A sub-filtration $\mathscr{G}_t = \sigma(p_1, \ldots, p_t)$ **(NO data-dependent stopping!)**

Fact: *No test martingale for this null has power under the data filtration.*

p_i: *conformal "p-values" that deviate from 1/2 under change*

What Goes Wrong When Combining E-Processes Across Filtrations?

For a fixed sample size n**:**

- Suppose that E_n and E'_n are two *arbitrary* e-values for H_0.
- Their **mean** is also an e-value for H_0.

$$\mathbb{E}_{H_0}\left[\frac{1}{2}\left(E_n + E'_n\right)\right] = \frac{1}{2}\left(\mathbb{E}_{H_0}[E_n] + \mathbb{E}_{H_0}[E'_n]\right) \le 1$$

At a ***data-dependent*** **stopping time** $\tau^{\mathbb{F}}$**:**

- Suppose that $(\mathfrak{e}_t)_{t\ge 0}$ and $(\mathfrak{e}'_t)_{t\ge 0}$ are two *arbitrary* e-processes for H_0.
- $\mathfrak{e}$ is defined in data filtration $\mathbb{F}$;
  $\mathfrak{e}'$ is defined in **a sub-filtration** $\mathbb{G} \subseteq \mathbb{F}$**.**

$$\mathbb{E}_{H_0}\left[\frac{1}{2}\left(\mathfrak{e}_{\tau^{\mathbb{F}}} + \mathfrak{e}'_{\tau^{\mathbb{F}}}\right)\right] = \frac{1}{2}\left(\mathbb{E}_{H_0}[\mathfrak{e}_{\tau^{\mathbb{F}}}] + \mathbb{E}_{H_0}[\mathfrak{e}'_{\tau^{\mathbb{F}}}]\right) \not\le 1$$

The General Question

Can we combine arbitrary e-processes **across filtrations**

such that the combined evidence is an e-process?

Combining E-Processes via $\mathfrak{e}$-Lifting

First Result: P-Processes Can Be Lifted "Freely"

- A $[0,1]$-valued process $(\mathfrak{p}_t)_{t\geq 0}$ is a **p-process** ("anytime-valid p-value") for H_0 defined in a filtration $\mathbb{F}$, if **for any $\mathbb{F}$-stopping time τ,** the random variable $\mathfrak{p}_\tau$ is a p-value for H_0.

Theorem ($\mathfrak{p}$-lifting). Let $(\mathfrak{p}_t)_{t\geq 0}$ be a p-process for H_0 **in a sub-filtration $\mathbb{G} \subseteq \mathbb{F}$.**

Then, $(\mathfrak{p}_t)_{t\geq 0}$ is a p-process for H_0 **in the original filtration $\mathbb{F}$.**

More generally, any "probability statement" translates to finer filtrations.

Main Result: Lifting E-Processes Using Adjusters

Recall that $(\mathfrak{e}_t)_{t\geq 0}$ is an **e-process for** H_0 in $\mathbb{F}$ if $\mathbb{E}_{H_0}\left[\mathfrak{e}_\tau\right] \leq 1$ for any $\mathbb{F}$-stopping time τ.

Theorem (e-lifting). Let $(\mathfrak{e}_t)_{t\geq 0}$ be an e-process for H_0 **in a sub-filtration $\mathbb{G} \subseteq \mathbb{F}$.**

For any adjuster A *(to be defined soon)*,

1. $\left(A(\mathfrak{e}_t)\right)_{t\geq 0}$ is an e-process for H_0 **in the data filtration $\mathbb{F}$.**
2. $\left(A\left(\mathfrak{e}_t^*\right)\right)_{t\geq 0}$ is an e-process for H_0 **in the data filtration $\mathbb{F}$.**

$(\mathfrak{e}_t^* = \max_{i\leq t}\mathfrak{e}_i)$

Proof Outline: $\mathfrak{e} \to \mathfrak{p} \to \mathfrak{e}^{\text{adj}}$

<u>Given:</u> An e-process $(\mathfrak{e}_t)_{t\geq 0}$ **in a sub-filtration $\mathbb{G} \subseteq \mathbb{F}$.**

1. Obtain a p-process $(\mathfrak{p}_t)_{t\geq 0}$ **in $\mathbb{G}$** (via Ville's inequality):

$$\mathfrak{p}_t = 1/\mathfrak{e}_t^*.$$

2. **By the $\mathfrak{p}$-lifting theorem,** $(\mathfrak{p}_t)_{t\geq 0}$ is also a p-process **in $\mathbb{F}$.**

3. Convert into an e-process $(\mathfrak{e}_t^{\text{adj}})_{t\geq 0}$ **in $\mathbb{F}$** via a p-to-e calibrator **C:**

$$\mathfrak{e}_t^{\text{adj}} = C(\mathfrak{p}_t).$$

"Adjustment"
(Dawid et al., 2011)

*What Are Adjusters?

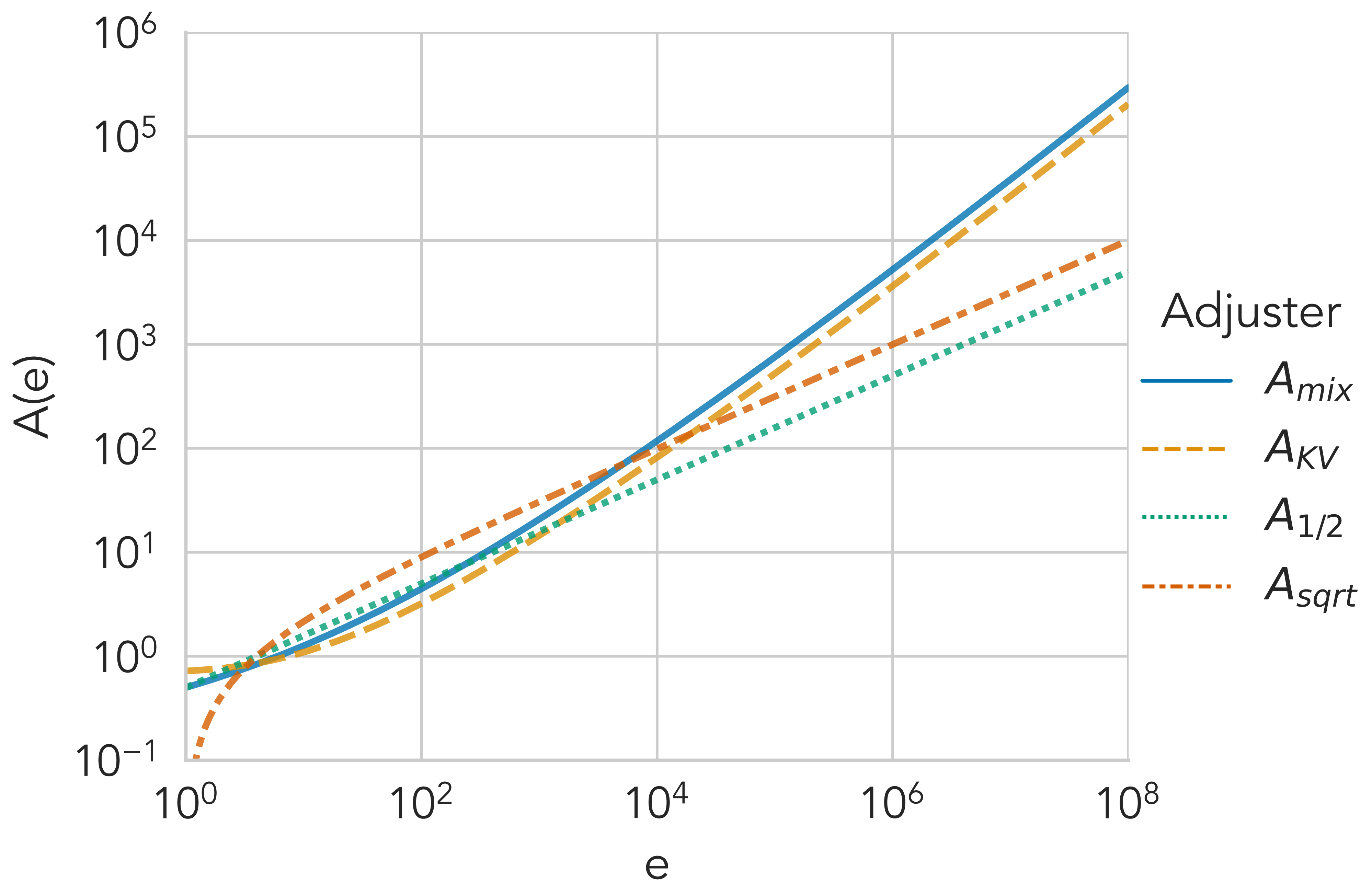
- Any increasing, right-continuous function $A : [1, \infty] \to [0, \infty]$ that satisfies:

$$\boxed{\int_1^\infty \frac{A(e)}{e^2} de = 1.}$$

- Recommended:

$$A_{\text{mix}}(e) = \frac{e - 1 - \log(e)}{\log^2(e)}.$$

(linear up to log terms)

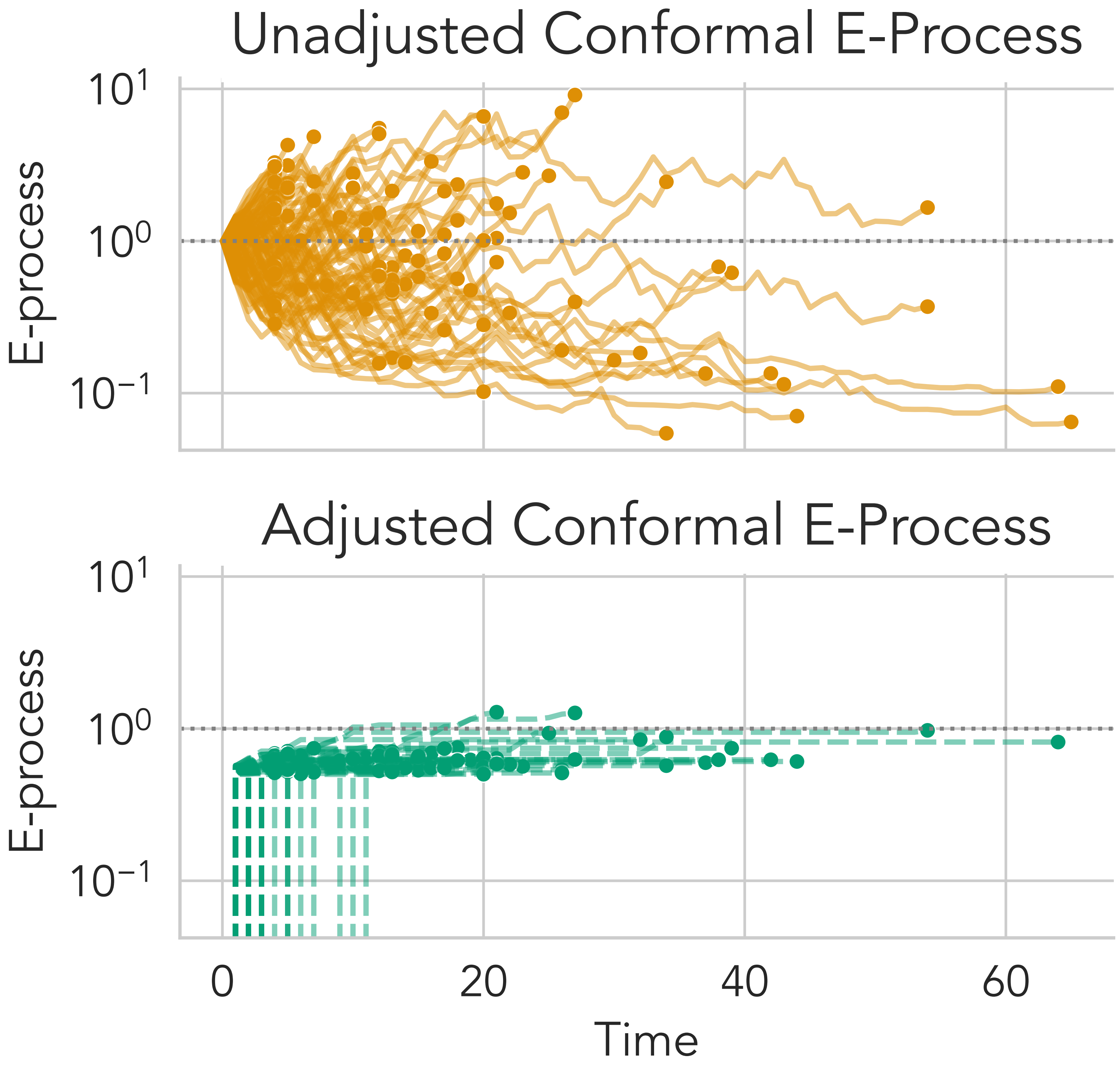


Dawid et al. (2011a;b), Shafer et al. (2011); Koolen & Vovk (2014)

Testing Randomness Online: Null Case

Data: i.i.d. Bernoulli.

$\tau^{\mathbb{F}}$ = **first time we observe five consecutive 0's (invalid in $\mathbb{G}$)**



$$\mathbb{E}_{H_0}\left[\tilde{\mathfrak{e}}_{\tau^{\mathbb{F}}}\right] \approx 1.33.$$

$$\mathbb{E}_{H_0}\left[A(\tilde{\mathfrak{e}}^*_{\tau^{\mathbb{F}}})\right] \approx 0.47.$$

The General Recipe: Adjust-Then-Combine

Given:

- A null hypothesis H_0.
- An e-process $(\mathfrak{e}_t)_{t \geq 0}$ for H_0 that is valid **in the data filtration $\mathbb{F}$.**
- Another e-process $(\tilde{\mathfrak{e}}_t)_{t \geq 0}$ for H_0 that is valid **only in a sub-filtration $\mathbb{G} \subseteq \mathbb{F}$.**
- An adjuster A.

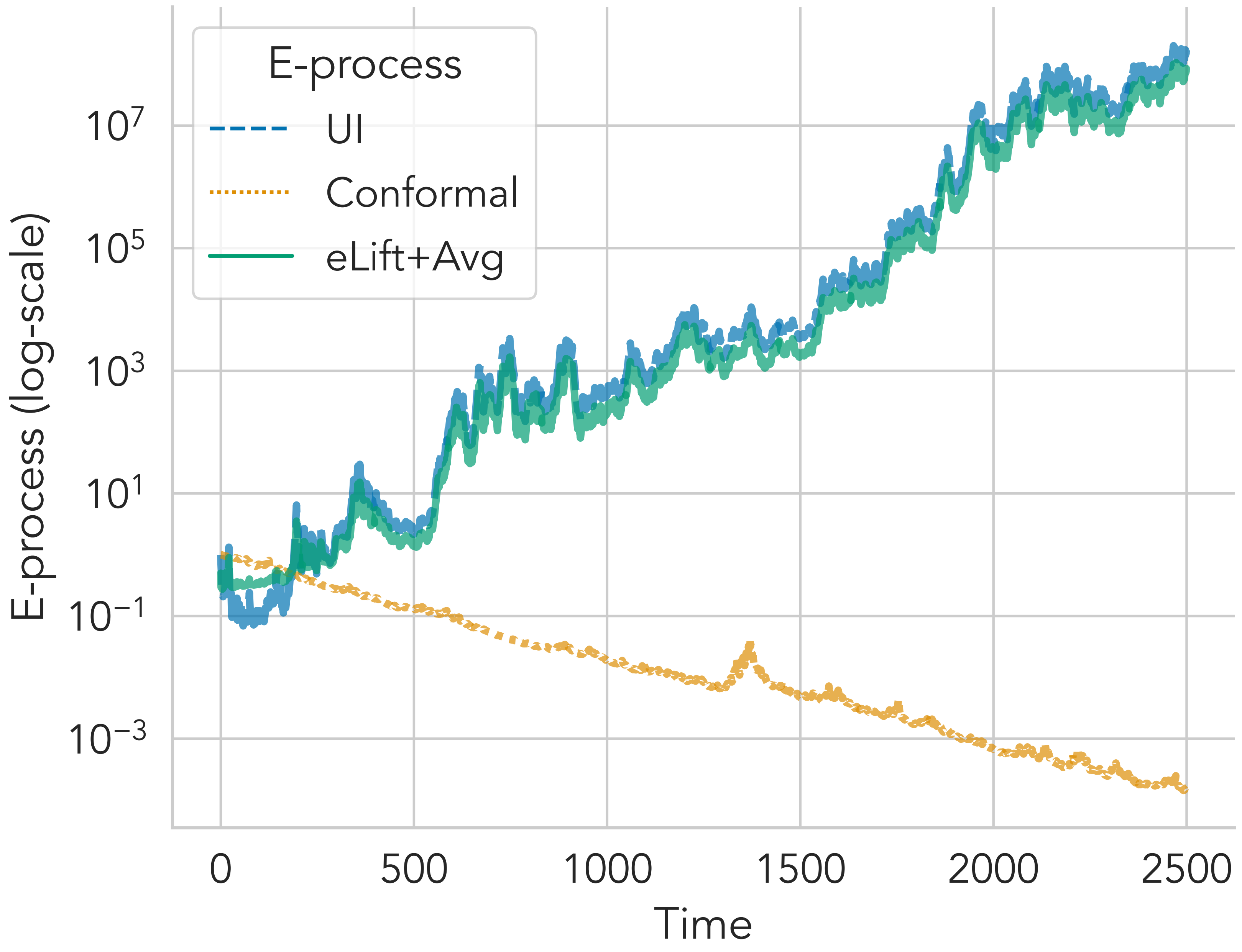
At any data-dependent stopping time $\tau^{\mathbb{F}}$:

1. Take the running maximum of $(\tilde{\mathfrak{e}}_t)_{t \geq 0}$: $\tilde{\mathfrak{e}}^*_\tau = \max_{i \leq \tau} \tilde{\mathfrak{e}}_i$.
2. **Adjust** that e-process: $A\left(\tilde{\mathfrak{e}}^*_\tau\right)$.
3. **Combine** them by averaging:

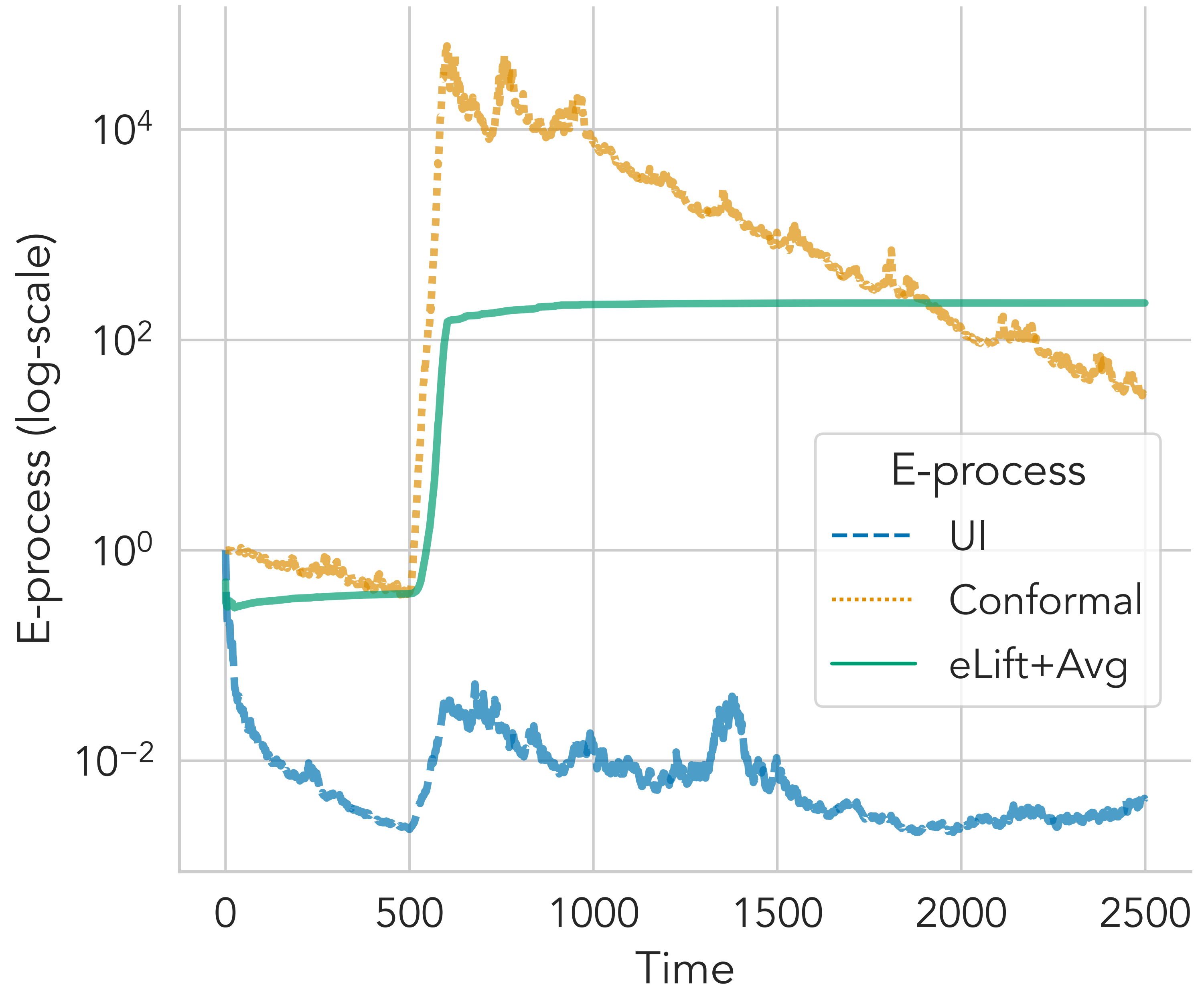
$$\bar{\mathfrak{e}}_\tau = \frac{1}{2}\left[\mathfrak{e}_\tau + A\left(\tilde{\mathfrak{e}}^*_\tau\right)\right].$$

Testing Randomness Online: Alternative Cases

Combined ("eLift+Avg"): $\bar{e}_t = \frac{1}{2}\left[e_t^{\mathrm{UI}} + A\left(\left(e_t^{\mathrm{conf}}\right)^*\right)\right]$



Alternative #1: First-order Markov



Alternative #2: Changepoint (@ T=500)

The combined e-process achieves power against both alternatives.

Additional Results & Discussion

Additional Results & Implications

1. **Applications to other sequential composite testing problems.**
   - *Evaluating/Comparing k-step-ahead forecasters*
   - *Independence testing; group-invariant null testing*
2. **In a formal sense, using an adjuster is necessary for lifting e-processes.**

Theorem (informal):

*Any deterministic & increasing function that maps $(\mathfrak{e}_t^*)_{t \geq 0}$ to an e-process (in the same filtration) is **necessarily** an adjuster.*

E-Process vs. P-Process: Contrasts & Synergies

- **Contrasts**

  1. Usually, we can easily combine arbitrary e-processes but **not** p-processes.
  2. On the other hand, $\mathfrak{p}$-lifting is free, but $\mathfrak{e}$-lifting is **not**.

- **Synergies**

  1. We can lift e-processes by **calibrating** them into p-processes (via adjusters).
  2. We can combine arbitrary e-/p-processes across arbitrary filtrations.

Future Work

1. **Sequential E-Multiple Testing**

   - *Adaptively stopping w.r.t. multiple e-processes can pose challenges!*

2. **Optimal Combination Strategies for E-Processes in Specific Scenarios**

   - *Are there alternative strategies that are more powerful in <u>specific</u> combination scenarios?*

   - *Is there a way to avoid taking the running maximum?*

MIND THE ~~GAP~~
FILTRATION

Thank You

For more, check out YJ's webpage:

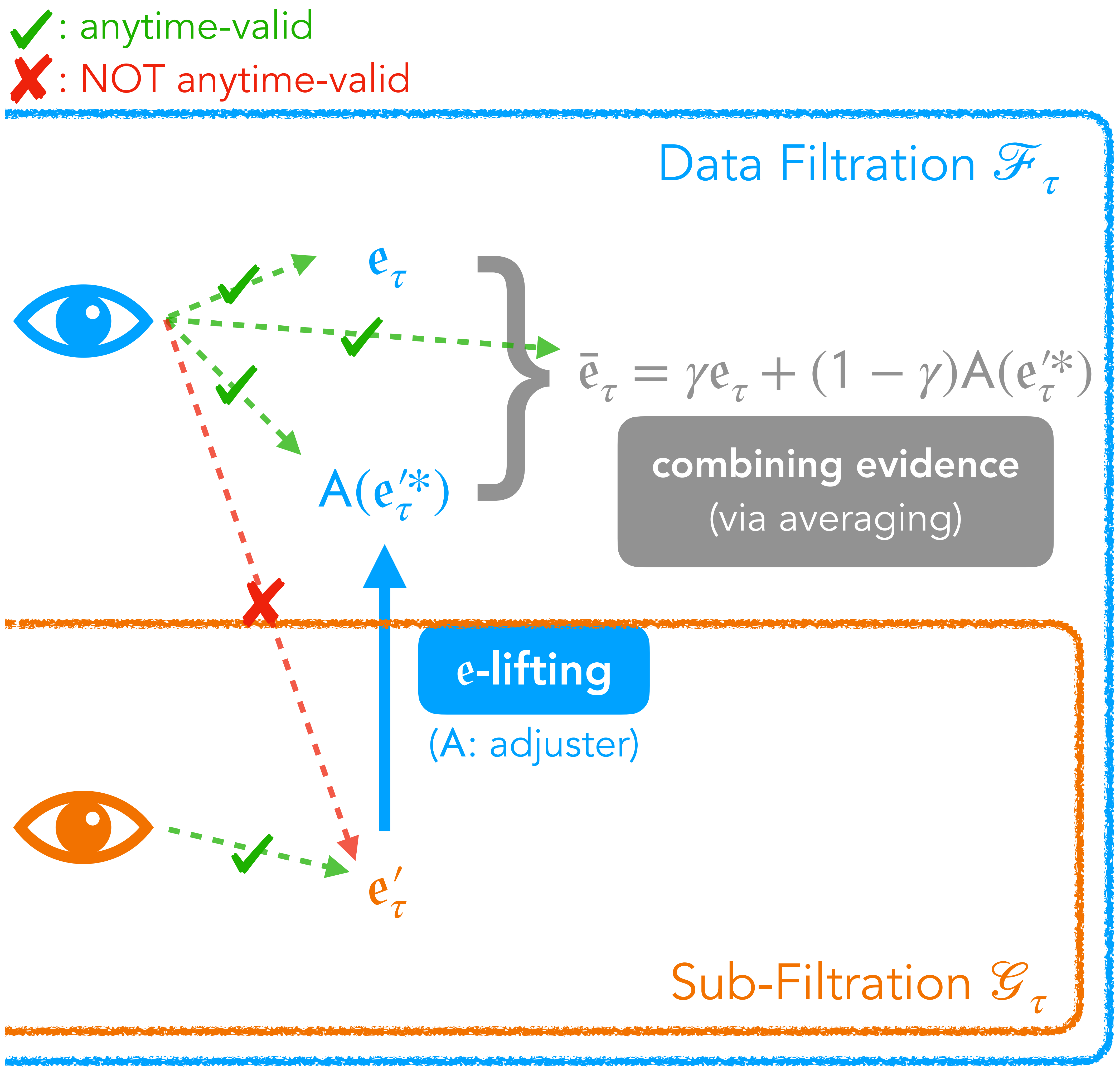
https://yjchoe.github.io/

Questions?

Yo Joong Choe & Aaditya Ramdas (2024).
"Combining Evidence Across Filtrations."
Preprint: https://arxiv.org/abs/2402.09698

Appendix

Combining evidence across filtrations via e-lifting



Testing-By-Betting

Protocol (Testing a probability by betting):

Casino proposes a **probability ("null hypothesis")** P over $\mathscr{Y}^\infty$.

Skeptic starts with initial wealth $M_0 = 1$.

For rounds $t = 1, 2, \ldots$:

1. Skeptic chooses a betting function $S_t : \mathscr{Y} \to \mathbb{R}_{\geq 0}$ such that $\mathbb{E}_P[S_t(Y_t)] = 1$.
2. Reality announces the outcome $y_t \in \mathscr{Y}$.
3. **Skeptic's wealth** is updated: $M_t = M_{t-1} \cdot S_t(y_t)$.

cf. Shafer (2021); Cournot (1800s)

cf. Shafer (2021); Cournot (1800s)

Testing-By-Betting

Bet against the null; accumulated wealth is the evidence against the null

Protocol (Testing a probability by betting).
Players: Casino, Skeptic, Reality

Casino proposes a **probability** P on $\mathcal{Y}^\infty$.
Skeptic starts with initial wealth $M_0 = 1$.
For rounds $t = 1, 2, \ldots$:

1. Skeptic chooses a betting function $S_t : \mathcal{Y} \to \mathbb{R}_{\geq 0}$ such that $\mathbb{E}_P[S_t(Y_t)] = 1$.
2. Reality announces the outcome $y_t \in \mathcal{Y}$.
3. Skeptic's wealth is updated as: $M_t = M_{t-1} \cdot S_t(y_t)$.

The Fundamental Principle of Testing-by-Betting:
Skeptic can discredit P *to the extent that* M_t *is large.*

Skeptic's wealth $(M_t)_{t\geq 0}$, a **test martingale** for P, is

- **Adapted:** at round t, Skeptic bets only knowing information up to round $t-1$.
- **Anytime-valid:** under P, Skeptic's expected wealth is bounded under *optional stopping*, i.e.,

For any stopping time τ, $\mathbb{E}_P[M_\tau] \leq 1$.

E-values generalize likelihood ratios

- Outside of a sequential setup (e.g. i.i.d. data and fixed sample size), we can still define "e-values". Given a probability distribution P, an **e-value** E is a nonnegative r.v. that satisfies

$$\boxed{\mathbb{E}_P[E] \le 1.}$$

- When testing a point null $H_0 : Y \sim P$ against a point alternative $H_1 : Y \sim Q$, the **likelihood ratio** Q/P is an e-value:

$$\mathbb{E}_P\left[\frac{Q(X)}{P(X)}\right] = \int \frac{Q(x)}{P(x)} P(x)dx = \int Q(x)dx = 1$$

- In the game-theoretic setup, the skeptic's bet in each round is an e-value. (The bet induces an "implied alternative" Q.)
- Any e-process at a stopping time is an e-value.

cf. Vovk and Wang (2021); Shafer (2021); Grünwald et al. (2024)

The Equivalence Lemma

Ramdas et al. (2020); Howard et al. (2021)

Let $(\xi_t)_{t\geq 1}$ be a sequence of events adapted to a filtration $\mathbb{G}$. (E.g., $\xi_t = \{\mathfrak{p}_t \leq \alpha\}$.)

Given any probability P and any $\alpha \in (0, 1)$, the following statements are equivalent:

(a) **Time-uniform validity**: $P\left(\cup_{t\geq 1} \xi_t \right) \leq \alpha$.

(b) **Random time validity**: for any (possibly infinite) random time T, $P(\xi_T) \leq \alpha$.

(c) **$\mathbb{G}$-anytime-validity**: for any (possibly infinite) $\mathbb{G}$-stopping time $\tau^{\mathbb{G}}$, $P(\xi_{\tau^{\mathbb{G}}}) \leq \alpha$.

$\mathfrak{p}$-lifting Follows Directly From "The Lifting Lemma"

Lemma. Let $(\xi_t)_{t\geq 1}$ be a sequence of events **adapted to a sub-filtration $\mathbb{G} \subseteq \mathbb{F}$.**

Given any probability P and any $\alpha \in (0, 1)$, the following statements are **equivalent**:

(a) $\mathbb{G}$-**anytime-validity**: for any $\mathbb{G}$-stopping time $\tau^{\mathbb{G}}$, $P(\xi_{\tau^{\mathbb{G}}}) \leq \alpha$.

(b) $\mathbb{F}$-**anytime-validity**: for any $\mathbb{F}$-stopping time $\tau^{\mathbb{F}}$, $P(\xi_{\tau^{\mathbb{F}}}) \leq \alpha$.

Any "probability statement" translates to finer filtrations.

Adjusters $\Longleftrightarrow$ P-to-E Calibrators

- A decreasing, left-continuous function $C : [0,1] \to [0,\infty]$ is a **(p-to-e) calibrator** if

$$\boxed{\int_0^1 C(\mathfrak{p})d\mathfrak{p} \leq 1.}$$

- It is *admissible* if the above holds with equality.

- There is a straightforward **1-to-1 correspondence between calibrators and adjusters.** Setting $A(\mathfrak{e}) = C(1/\mathfrak{e})$, and by change-of-variables ($\mathfrak{p} = 1/\mathfrak{e}$),

$$\int_1^\infty \frac{A(\mathfrak{e})}{\mathfrak{e}^2}d\mathfrak{e} = \int_1^\infty \frac{C(1/\mathfrak{e})}{\mathfrak{e}^2}d\mathfrak{e} = \int_0^1 C(\mathfrak{p})d\mathfrak{p} \underset{(=)}{\leq} 1.$$

cf. Shafer et al. (2011); Vovk & Wang (2021)

Other Examples in the Literature

1. **Multi-step forecast evaluation/comparison**

   - A valid strategy is to construct e-processes $(e_t^{[k]})_{t \geq 0}$ **in *different coarsenings of the data filtration*, say** $\mathbb{G}^{[k]} \subsetneq \mathbb{F}$**.** (Henzi & Ziegel, 2022)

   - To evaluate across all coarsened filtrations, we need to $\mathfrak{e}$-lift all h e-processes!

2. **Sequential independence testing**

   - For this problem, there is no nontrivial test martingale w.r.t. the data filtration. (Henzi & Law, 2024) Existing e-processes thus operate on different coarsened filtrations.

[1]Henzi & Ziegel (2022); Arnold et al., (2023); Choe & Ramdas (2023)

[2]Balasubramani & Ramdas (2016); Shekhar & Ramdas (2023); Podkopaev et al. (2023); Henzi & Law (2024)

Example: Comparing Multi-Step Sequential Forecasters

- Suppose we compare two sequential forecasters with lag h using some scoring rule S w.r.t. $\mathbb{F} = (\mathscr{F}_t)_{t\geq 0}$:

$$\Delta_t^{[k]} = \frac{1}{|I_t^{[k]}|} \sum_{i \in I_t^{[k]}} \mathbb{E}\left[S(p_i, y_{i+h-1}) - S(q_i, y_{i+h-1}) \mid \mathscr{F}_{i-1}\right], \quad \forall k \in [h].$$

  - If $h = 2$, $\Delta_t^{[0]}/\Delta_t^{[1]}$ measures the **average forecast score difference** on **even/odd** days.

- When testing for the null $\mathscr{H}_0^{[k]} : \Delta_t^{[k]} \leq 0, \forall t$, for each offset k, we need to construct an e-process $(\mathfrak{e}_t^{[k]})_{t\geq 0}$ **under different coarsening of the filtration** $\mathbb{F}$ **for each** k (updates on every even/odd days).

Each $(\mathfrak{e}_t^{[k]})_{t\geq 0}$ is an e-process for $\mathscr{H}_0^{[k]}$, but only w.r.t. the sub-filtration $\mathbb{G}^{[k]} \subsetneq \mathbb{F}$.

- To test for **the combined null** $\mathscr{H}_0 : \Delta_t^{[k]} \leq 0, \forall t, \forall k$ (an intersection), we want to **$\mathfrak{e}$-lift** all h e-processes into the data filtration $\mathbb{F}$ before combining them:

$$\bar{\mathfrak{e}}_t = \frac{1}{h} \sum_{k=1}^{h} A\left((\mathfrak{e}_t^{[k]})^*\right), \forall t.$$

Henzi & Ziegel (2022)
Arnold et al. (2022)
Choe & Ramdas (2023)

Example: Testing Independence

- Given an i.i.d. stream of paired data $Z_t = (X_t, Y_t) \sim P_{XY}$, suppose we test if the joint distribution factorizes:

$$\mathscr{H}_0 : P_{XY} = P_X \times P_Y \quad \text{vs.} \quad \mathscr{H}_1 : P_{XY} \neq P_X \times P_Y.$$

- Similar to the exchangeability null, there exist no nontrivial test martingale adapted to the data filtration $\mathbb{F}$. Two known e-processes include:
    - **Pairwise betting** (SR'23; PBKR'23; SR'24): adapted to the filtration w/ pairs of data.
    - **Rank-based test martingale** (HL'23): adapted to the filtration w/ rank stats of data.
- In this case, BOTH e-processes are constructed w.r.t. their own, non-overlapping sub-filtrations. So we should lift both of them before taking the average.

cf. Balasubramani & Ramdas (2016); Shekhar & Ramdas (2023); Podkopaev et al. (2023); Henzi & Law (2023)

Theorem: Equivalent characterizations of adjusters

Game-Theoretic Definition

There is a $\mathbb{G}$-safe NSM that dominates $A(e_t^*)$ for any $\mathbb{G}$-safe e-process $(e_t)_{t \geq 0}$.

Equivalent defn. for e-processes

E-Process Definition

"$A(e_t^*)$ is a $\mathbb{G}$-safe e-process for any e_t"

$\mathbb{E}_P[A(e_\tau^*)] \leq 1$ for any $\mathbb{G}$-stopping time τ and any $\mathbb{G}$-safe e-process $(e_t)_{t \geq 0}$.

Dawid et al. (2011)

e-lifting

Trivial

$$\int_1^\infty \frac{A(y)}{y^2} dy \leq 1.$$

$\mathbb{E}_P[A(e_\tau^*)] \leq 1$ for any $\mathbb{F}$-stopping time τ and any $\mathbb{G}$-safe e-process $(e_t)_{t \geq 0}$.

cf. Dawid et al. (2011a;b); Shafer et al. (2011); Koolen & Vovk (2014)

A Corollary on Coarsening the Filtration

Corollary. Let $\mathscr{P}$ be a composite null and let $\mathscr{Q}$ be a composite alternative.

Suppose there exists a $\mathscr{Q}$-powerful* e-process for $\mathscr{P}$ in a sub-filtration $\mathbb{G}$ of $\mathbb{F}$.

Then, there exists a $\mathscr{Q}$-powerful e-process for $\mathscr{P}$ in $\mathbb{F}$.

*An e-process for $\mathscr{P}$ is $\mathscr{Q}$-**powerful** if, for any $Q \in \mathscr{Q} \setminus \mathscr{P}$, $\limsup_{t\to\infty} \mathfrak{e}_t = \infty$, Q-almost surely.

- *Interestingly, this is **NOT** the case if "e-process" is replaced with "test martingale".*

Is it *necessary* to adjust the e-process?

- Suppose I claim to have a function that, given any composite null, if you give me *any* e-process for the null a coarse filtration, then the function can transform it into an e-process for the same null in the data filtration.

- **Is the function necessarily an adjuster?**

Necessity of Adjusters for $\mathfrak{e}$-lifting

<u>Theorem.</u> Let $A : [1, \infty] \to [0, \infty]$ be an increasing function. The following are **equivalent**:

(a) A is an adjuster.

(b) A is an **"$\mathfrak{e}$-lifter"**: given any $\mathscr{P}$, for any e-process $(\mathfrak{e}_t)_{t \geq 0}$ for $\mathscr{P}$ **in** $\mathbb{G}$ and **for *any* finer filtration** $\mathbb{F} \supseteq \mathbb{G}$**,** $\left(A(\mathfrak{e}_t^*)\right)_{t \geq 0}$ is an e-process for $\mathscr{P}$ **in** $\mathbb{F}$.

In particular, any deterministic & increasing function that maps

$\max_{i \leq t} \mathfrak{e}_i$ ***to some*** $\mathfrak{e}_t'$ ***(for each*** t***) is necessarily an adjuster.***

A Characterization Theorem for Adjusters

<u>Theorem.</u> Let $A : [1, \infty] \to [0, \infty]$ be an increasing function. The following are <u>equivalent</u>:

(a) A is an adjuster, i.e., it satisfies $\displaystyle\int_1^\infty \frac{A(\mathfrak{e})}{\mathfrak{e}^2} d\mathfrak{e} \le 1.$

(b) A is an "adjuster for test supermartingales" (previous slide).

(c) A is an **"adjuster for e-processes"**: given any $\mathscr{P}$, for any e-process $(\mathfrak{e}_t)_{t\ge0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$, there exists another e-process $(\mathfrak{e}'_t)_{t\ge0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$ such that, for all t, $A(\mathfrak{e}^*_t) \le \mathfrak{e}'_t$.

(d) A is an **"$\mathfrak{e}$-lifter"**: given any $\mathscr{P}$, for any e-process $(\mathfrak{e}_t)_{t\ge0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$, **and *any* finer filtration $\mathbb{F} \supseteq \mathbb{G}$,** $(A(\mathfrak{e}^*_t))_{t\ge0}$ is an e-process for $\mathscr{P}$ **w.r.t. $\mathbb{F}$**.

(e) Given any $\mathscr{P}$, for any e-process $(\mathfrak{e}_t)_{t\ge0}$ for $\mathscr{P}$ w.r.t. $\mathbb{G}$, $(A(\mathfrak{e}^*_t))_{t\ge0}$ is an e-process for $\mathscr{P}$ w.r.t. $\mathbb{G}$.

A game-theoretic definition of adjusters

How can we make betting on the running maximum a "fair game"?

- An increasing function A is an adjuster **if and only if**, for every test supermartingale $(M_t)_{t \geq 0}$ for some P, there exists a test supermartingale $(M'_t)_{t \geq 0}$ for P s.t.

$$\boxed{A(M^*_t) \leq M'_t, \quad \forall t.}$$

A is an "adjuster for test supermartingales"

- Game-theoretically, adjusters allow betting with the **running maximum** of the gambler's wealth.
  - A is an adjuster if and only if, in Protocol 1, Rival Skeptic has a betting strategy to ensure that

$$A(\mathscr{K}^*_t) \leq \mathscr{K}'_t.$$

Protocol 1 Competitive scepticism

$\mathcal{K}_0 := 1$ and $\mathcal{K}'_0 := 1$
for $n = 1, 2, \ldots$ **do**
  Forecaster announces $\mathcal{E}_n \in \mathbf{E}$
  Sceptic announces $f_n \in [0, \infty]^{\mathcal{X}}$ such that $\mathcal{E}_n(f_n) \leq \mathcal{K}_{n-1}$
  Rival Sceptic announces $f'_n \in [0, \infty]^{\mathcal{X}}$ such that $\mathcal{E}_n(f'_n) \leq \mathcal{K}'_{n-1}$
  Reality announces $x_n \in \mathcal{X}$
  $\mathcal{K}_n := f_n(x_n)$ and $\mathcal{K}'_n := f'_n(x_n)$
end for

Dawid et al. (2011a;b); Shafer et al. (2011)

Comparing k-Step-Ahead Weather Forecasters ☔

- **Data:** Precipitation data at four airport locations (Brussels, Frankfurt, London, & Zurich), 2007—2017. *(Source: the European Centre of Medium-Range Weather Forecasts)*
- **Forecasting Task:** Using the 2007—2012 data, make accurate probability forecasts for 2012—2017.
- **Forecasting Methods:**
  - *Method #1: Isotonic Distributional Regression (IDR) Ensemble*
  - *Method #2: Heteroskedastic Censored Logistic Regression (HCLR) Ensemble*
  - *Baseline: Climatology (i.e., historical mean)*
- **Evaluation:** Mean expected Brier score difference

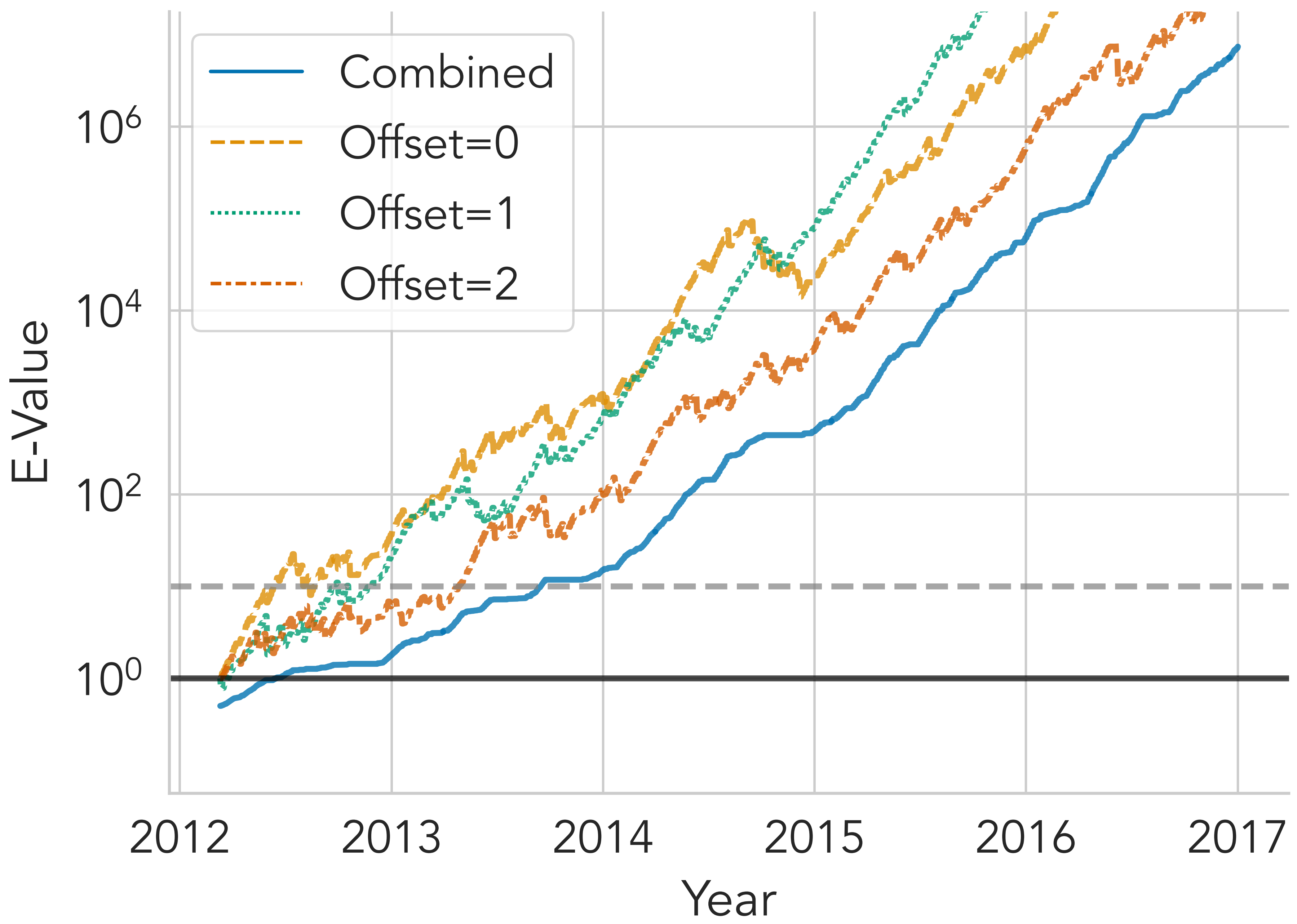
Vannitsem et al. (2018); Henzi et al. (2021)

Data: Precipitation in Zurich Airport

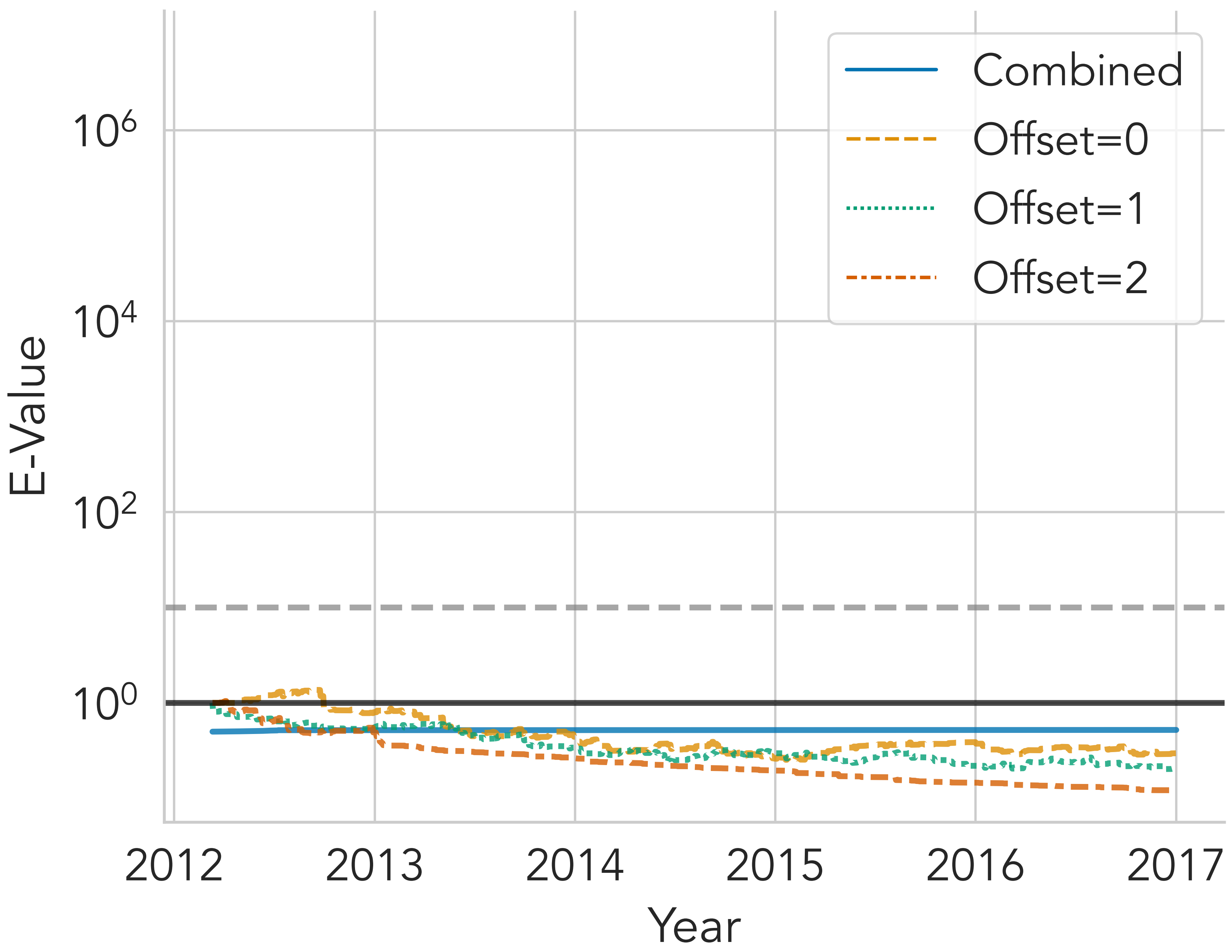
Comparing 3-Day-Ahead Weather Forecasters ☔

**<u>Note:</u> Only the "Combined" version is valid at data-dependent sample sizes*

Comparison #1: IDR vs. Climatology



Comparison #2: IDR vs. HCLR



There is strong evidence to discredit Climatology over IDR.
(passes the baseline)

There isn't enough evidence to discredit HCLR over IDR.
(consistent with prior findings)

End of Slides