BACKTESTING RISK MEASURES

In particular Value at Risk and Expected Shortfall

Carlo Acerbi, Balazs Szekely

September 2016

MSCI

AGENDA

- What does backtesting mean?
- The Basel VaR vs ES dilemma
  - And the importance of backtestability
- VaR and ES: a closer look
  - Getting an intuition of why VaR is backtestable and ES (maybe) not
- Elicitability and backtestability
  - Formalizing intuition
- L'ES est mort, vive l'ES
  - How to backtest ES, nonetheless
- Did VaR create ES?
  - Indirect backtestability

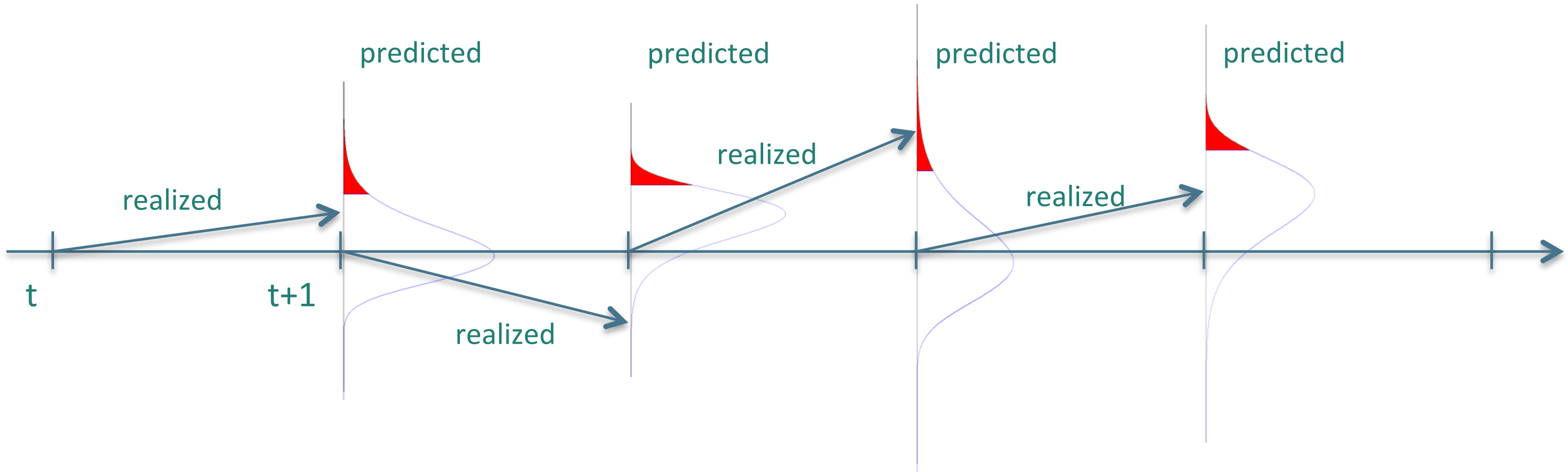
WHAT DOES BACKTESTING MEAN?

The Importance of Model Validation

TESTING PREDICTIONS

- **Setting**: a sequence of predictions on future events
- **Objective**: testing 'ex post' how good your predictions have been
- Ex: every week you forecast the winning horse for next Sunday's races
  - At the end of the season you can easily score your 'prediction model'
  - Why? Because the winning horse is publicly declared every week
- Ex: Banks have to forecast their risks and allocate capital accordingly
  - Risk models need to be validated -> backtested
  - But risk is not declared ex post
    - Is it always possible to backtest a risk measure?
    - What makes a measure backtestable or not?

WHAT DOES BACKTESTING MEAN?



- Backtesting a risk measure means testing forecasts against realizations. However
  - Distributions and risk measures do not materialize ex post like winning horses
  - Only one scenario at a time is revealed

- A risk measure is commonly said *backtestable* if there exists an observable test statistic, that allows to say whether predictions are over/under-estimated
  - E.g.: risk measure = VaR; test variable = VaR breaches counting

- Not all risk measures appear to be backtestable
  - However, a formal definition of 'backtestable' has long been missing

MSCI

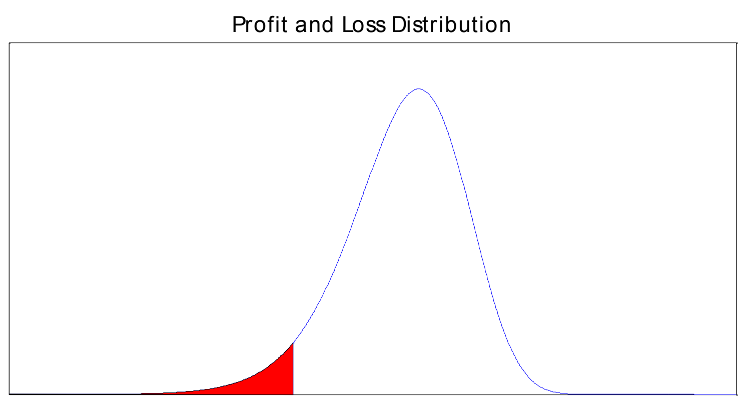
THE BASEL VAR - ES DILEMMA

And the Importance of Backtesting

MSCI

VAR AND ES

- VaR: the best of worst x% losses; → threshold of x% losses
- ES: the average of worst x% losses → expected x% loss



- ES
  - Multiple advantages: tail sensitivity, subadditivity (→ coherent), mathematical tractability, uniqueness, uses same risk models …
  - One drawback: how to backtest ES is still an open issue

BASEL REGULATION: VAR OR ES?

- 1994: RiskMetrics Technical Document popularizes "Value at Risk" (VaR)
- 1996: Basel Committee internal-based approach to capital adequacy, based on VaR
- 1997: Artzner et al. "Coherent Measures of Risk": axioms for sensible risk measures. VaR criticized for not complying
- 2001: Rockafellar and Uryasev, Acerbi and Tasche, define "Expected Shortfall" (ES, aka CVaR), a coherent measure of risk
- 2000s VaR and ES are widely adopted by financial institutions as complementary tools
- 2013/2016: Basel Committee replaces $VaR_{1\%}$ with $ES_{2.5\%}$
  - VaR is maintained for model backtesting

MSCI

BASEL VAR BACKTEST: TRAFFIC LIGHT SYSTEM

Table 2

Zone	Number of exceptions	Increase in scaling factor	Cumulative probability
Green zone	0	0.00	8.11%
	1	0.00	28.58%
	2	0.00	54.32%
	3	0.00	75.81%
	4	0.00	89.22%
Yellow zone	5	0.40	95.88%
	6	0.50	98.63%
	7	0.65	99.60%
	8	0.75	99.89%
	9	0.85	99.97%
Red zone	10 or more	1.00	99.99%

- (Source: BCBS FRTB 2013) Same mechanism since 1996

CURRENT DEBATE

- VaR has serious shortcomings but allows for immediate backtesting

- Is it possible to backtest ES?
  – If not, can regulation be based on models that can't be validated?

- Backtesting means **model validation**
  – Fundamental for a risk system, non negotiable feature
    - Only way to say if your model is doing a good job
  – Key property for a regulatory standard
    - ES now adopted by "Basel 3" (in FRTB drafts, since 2013)
      - IAIS reached the opposite conclusion for ICS

- Backtesting also <u>means nothing</u>, until someone gives a formal definition
  – We attempt a definition and try to put some order

VAR AND ES

A closer look

MSCI

VAR DEFINITION AND BACKTESTING

- VaR can be defined as

  $VaR_{\alpha} = -F^{-1}(\alpha)$

$Pr[X < -VaR_{\alpha}] = \alpha$

- The probability for a loss to exceed VaR_{α} is exactly α
  - In a sequence of N independent correct predictions, the losses exceeding VaR_{α} are **binomial distributed with mean $N\alpha$**
    - **More (less) exceptions show under- (over-) estimation of risk**

- Notice the expression $\mathbb{E}[\alpha - (X + VaR_{\alpha} < 0)] = 0$
  - A null expected value involving only VaR and X

ES DEFINITIONS AND BACKTESTING DIFFICULTY

- Several equivalent ES definitions
  - $ES_\alpha = \int_0^\alpha VaR_p \, dp$
  - $ES_\alpha = -\frac{1}{\alpha} \mathbb{E}[X(X+VaR_\alpha<0) - VaR_\alpha (\alpha - (X+VaR_\alpha<0))]$

    $= VaR_\alpha - \frac{1}{\alpha} \mathbb{E}[(X+VaR_\alpha)(X+VaR_\alpha<0)]$
    - Reduces to $ES_\alpha = -\mathbb{E}[X \mid X<-VaR_\alpha]$ only for continuous cdf
  - $ES_\alpha = \min_v [v - \frac{1}{\alpha} \mathbb{E}[(X+v)(X+v<0)]]$ (Uryasev and Rockafellar)
    - $VaR_\alpha = \operatorname{argmin}_v [\cdots]$

- Notice that there's no way to build an expression like

$\mathbb{E}[f(ES_\alpha, X)] = 0$

– f will also depend on VaR_α

- We will see that basically ES is not backtestable for this simple reason

ELICITABILITY AND BACKTESTABILITY

Formalizing why a measure is backtestable or not

ELICITABILITY

- A statistic $Y(X)$ is said to be **elicitable** if it solves

$$Y(X) = \arg\min_y \mathbb{E}[S(y,X)]$$

  for some scoring function $S(y,x)$

- Popular examples:
  - Mean: $X = \arg\min_y \mathbb{E}[(y-X)^2]$
  - Median: $X = \arg\min_y \mathbb{E}[|y-X|]$
  - α-quantile: $q_\alpha = \arg\min_y \mathbb{E}[(X-y)(\alpha-(X-y<0))]$ → VaR !

- If a statistic Y is elicitable, given forecasts y_t and realizations x_t the mean score ranks models: the lower, the better

$$S = \frac{1}{T}\sum_{t=1}^{T} S(y_t, x_t)$$

MODEL VALIDATION VS MODEL SELECTION

- If a measure is elicitable, we can rank models by their mean score
  – relative, not absolute scale
  – A single mean score value, per se, tells nothing
- A mean score compares different predictions to the same process
  – Ex: Bank A wants to select the best in a class of VaR forecast models
  – Elicitability → Model selection: relative scale
- Backtesting requires absolute significance
  – Ex: Bank A wants to validate its model
  – Backtesting → Model validation: absolute scale

IDENTIFIABILITY

- A statistic $Y(X)$ is said to be **identifiable** if it solves

$$\mathbb{E}[I(y,X)]=0 \quad \text{when} \quad y=Y(X)$$

  for some identification function $I(y,x)$

- Popular examples:
  - Mean: $\mathbb{E}[X-X]=0$
  - Median: $\mathbb{E}[1/2-(X-X<0)]=0$ (cont. cdf's)
  - α-quantile: $\mathbb{E}[\alpha-(X-q_{\alpha}<0)]=0$ (cont. cdf's)

- Under regularity conditions, elicitability and identifiability coincide

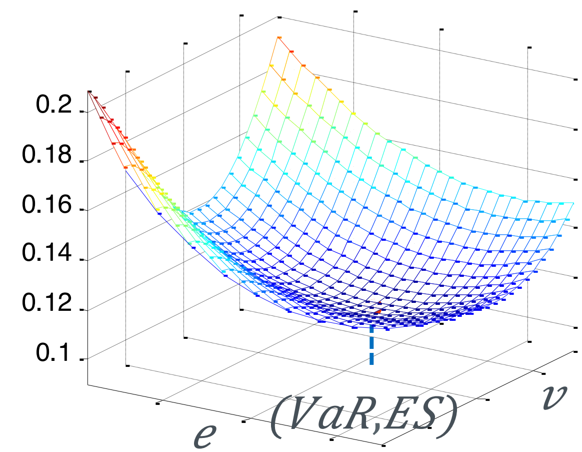
$$S(y,x)=\int^{y} I(t,x)\,dt \qquad \text{(Osband's principle, 1985)}$$

- Notice that VaR is always elicitable, but identifiable only if the cdf is continuous

MSCI

ES NEEDS VAR

- ES is neither elicitable nor identifiable per se (Gneiting, Ziegler, Weber, Bellini, …)
  - The pair (ES, VaR) however is both



- 2-Elicitability
  - Acerbi, Szekely 2014, $S^W(v,e,x)$
  - Ziegel, Fissler 2015, larger class of $S(v,e,x)$'s
- 2-Identification
  - Precise definition in Ziegel, Fissler 2015
  - From $ES_\alpha = VaR_\alpha - 1/\alpha \, \mathbb{E}[(X+VaR_\alpha)(X+VaR_\alpha<0)]$
    - $I_2(v,e,x) = \alpha(e-v) + (x+v)(x+v<0)$
- Striking parallel: also 'variance needs the mean', for the same properties

MSCI

BACKTESTABILITY – OUR DEFINITION

- **Definition**: a statistic $Y(X)$ is **backtestable** if there exists a **test function** $Z(y,x)$ such that

  $y \mapsto \mathbb{E}[Z(y,X)]$ is increasing

  $\mathbb{E}[Z(y,X)] \lesseqgtr 0$ when $y \lesseqgtr Y(X)$

- Very similar to identifiability, but not quite the same
  - We require more: monotonicity
  - We require less: we do not require $\mathbb{E}[Z(y,X)]=0$ in $y=Y(X)$

- No need for regularity conditions. VaR is always backtestable despite not always identifiable

MSCI

BACKTESTABLE RISK MEASURES

- It is immediate from $S(y,x)=\int^{y} Z(t,x)\,dt$ that
  - $Y(X)$ is backtestable iff it's elicitable with a y-convex scoring function S
    - The test function Z is the y-subdifferential of S
- All popular basic scoring functions happen to be already convex
  - VaR, Mean, Median, are backtestable
  - Expectiles are backtestable (Pandora's box)
- ES is not backtestable, because it's not elicitable
  - End of a long controversial debate

MSCI

HYPOTHESIS TESTING

- A **test statistic** of a backtestable risk measure is defined as

$$Z(y,x) = \frac{1}{T}\sum_{1}^{T} Z(y_t, x_t)$$

  - The H_0 (H_1) distribution of Z can always be resampled from predictive (resp. misspecified) distributions
  - Requires storage of predictive distributions day by day

- VaR is "super-backtestable" because the distribution of Z is model-independent, namely a binomial.
  - No need for resampling
  - Arguably a unique case (up to monotonic transformations)

MSCI

L’ES EST MORT, VIVE L’ES

How to backtest ES, nevertheless

HOW TO BACKTEST ES, NEVERTHELESS

- ES is 2-backtestable jointly with VaR
  - Follows immediately from its 2-identifiability
- This means that you can backtest ES only if you have previously backtested VaR, bilaterally, with success
- Generally speaking this seems a poor strategy
  - Hypothesis testing can rule out gross mistakes in VaR predictions, but never confirm that they are spot on!
- Questions
  - How sensitive is ES backtesting to VaR discrepancies?
  - How does a wrong VaR impact on the quality of ES backtest?

BACK TO BASICS

- We recall the expression valid for general distributions (Acerbi and Tasche 2001)

$$ES_{\alpha} = VaR_{\alpha} - \frac{1}{\alpha}\,\mathbb{E}[(X+VaR_{\alpha})(X+VaR_{\alpha}<0)]$$

- This yields a 2-identification/test function for VaR and ES

$$Z(v,e,x) = \alpha(e-v)+(x+v)(x+v<0)$$

- How sensitive is this function to VaR predictions?

BACK TO BASICS

- Now recall also Uryasev and Rockafellar's (2001) classical result

$$ES_{\alpha} = \min_{v} \left[v - 1/\alpha \, \mathbb{E}[(X+v)(X+v<0)] \right]$$

$$VaR_{\alpha} = \operatorname{argmin}_{v} [\cdots]$$

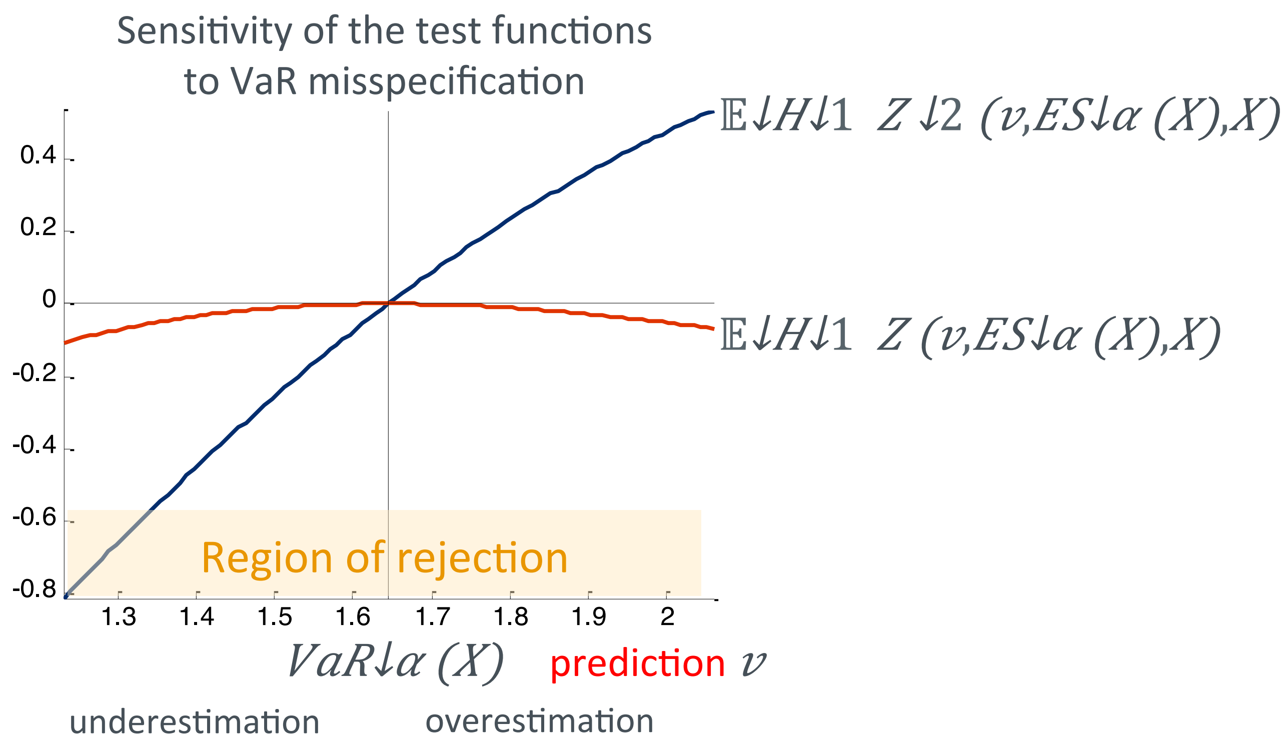
- Important consequence: if we backtest ES with $Z(v,e,x)$ then <u>any possible error in VaR prediction</u>
  1. Would make the backtest always <u>more prudential</u>
  2. Would have <u>no relevance at first order</u>

NEW PROPOSED TEST FOR ES

- $Z(v,e,x)=\alpha(e-v)+(x+v)(x+v<0)$
- Define $Z(\vec{X})=\sum_{t=1}^{T} \frac{Z(v_t, e_t, x_t)}{T\alpha e_t}$
- Hypotheses
  - $H_0: P_t^{[\alpha]} = F_t^{[\alpha]}$ *for all* t
  - $H_1: e_t \leq ES_{\alpha,t}^{F}$ for all t and $<$ for some t
    $v_t \sim VaR_{\alpha,t}^{F}$
- We have: $\mathbb{E}_{H_0}[Z]=0$ and $\mathbb{E}_{H_1}[Z]<0$
- VaR has to be tested separately, on two-sides, but loosely

SENSITIVITY TO VAR PREDICTION

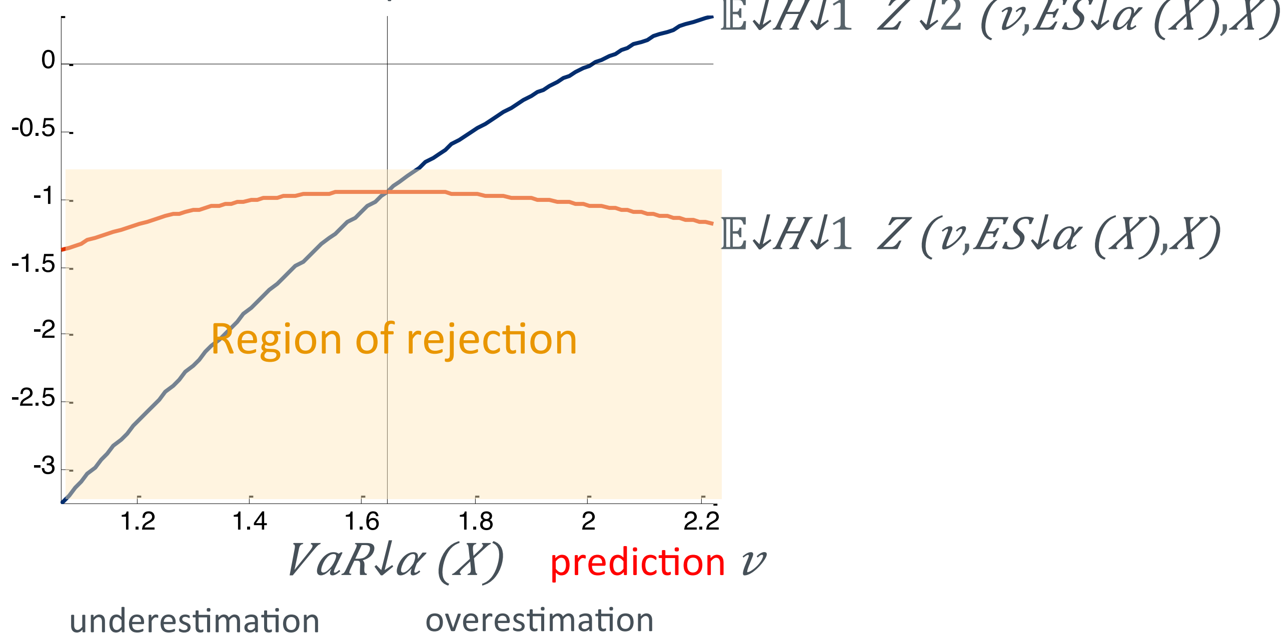
- Example: $\alpha=5\%$, true distributions are 250 i.i.d. N(0,1) returns
- **ES is estimated correctly**: $e=ES_\alpha(X)$



SENSITIVITY TO VAR PREDICTION

- Example: $\alpha = 5\%$, true distributions are 250 i.i.d. N(0,1) returns
- **ES is underestimated**: $e = ES_{\alpha}(X) - 1$

Sensitivity of the test functions
to VaR misspecification



CONSEQUENCES, IN PRACTICE

- The new test for ES lends itself to widespread adoption
  - It requires joint but mild validation of VaR
  - At worst, leads to a more penalizing test, in the case when VaR predictions should be completely wrong

- The test of ES requires nonetheless the day-by-day storage of predictive distributions for bootstrapping the $H\downarrow 0$ distribution of the test statistic and the computation of a p-value
  - VaR backtesting on the contrary requires storage of only one number per day, the prediction

- The method can be adopted for a traffic light system: it could be adopted in Basel regulation

DID VAR CREATE ES?

Indirect backtestability

STRIKING PARALLEL WITH VARIANCE

- Remember that also variance is 2-identifiable with the mean because of $\sigma^2 = \mathbb{E}[(X - X)^2]$

- Remember now that also in the case of variance we can express it as

$$\sigma^2 = \min_m \mathbb{E}[(X - m)^2]$$

- Exact same result on the 1st order irrelevance of misspecifications of the mean when backtesting the variance

- Coincidence?

- As a matter of fact, people have backtested variance for decades, without worrying much is the mean prediction was spot on

NOT A COINCIDENCE

- Suppose Y_1 backtestable: $Y_1(X) = \arg\min_y \mathbb{E}[S(y,X)]$

- Then $Y_2(X) = \min_y \mathbb{E}[S(y,X)] = \mathbb{E}[S(Y_1(X),X)]$ is automatically 2-backtestable with Y_1 from this very equation!
  - And tests of Y_2 based on this will be insensitive at 1^{st} order to Y_1 predictions and always biased in one direction only

- The couples (mean, variance) and (VaR, ES) are just two examples of this mechanism
  - Measures like Variance and ES are *indirectly backtestable*

- We can say that ES is generated by the backtestability of VaR!
  - VaR created its enemy

MSCI

CONCLUSIONS

CONCLUSIONS 1/2

- We provide a natural definition of backtestability which turns out to coincide with convex elicitability
- Most popular elicitable statistics happen to be convex too, hence backtestable. All non-elicitable statistics, notably ES, are not backtestable
- However the natural 2-identification function for VaR and ES has remarkable features
  - ES is 2-backtestable with VaR and the dependence of the ES test on VaR is zero at 1^{st} order and always of prudential type
  - We introduce a new test of ES based on these results that opens the way for valid practical backtests for ES

MSCI

CONCLUSIONS 2/2

- We notice that 2-backtestability of ES (and the above properties) is just an instance of a general situation which we term indirect backtestability, where a statistics is generated by the backtestability of another statistics.
  - Mean and variance are in the same exact relationship as VaR and ES
  - Variance and ES are just two instances of indirect backtestability

MODEL RISK KILLED MODELS

A short interlude

MSCI

BASEL TOURNAMENT

MSCI

AND THE WINNER IS ….

MSCI

THANKS FOR NOT FALLING ASLEEP

MSCI

CONTACT US

AMERICAS

Americas	1 888 588 4567 *
Atlanta	+ 1 404 551 3212
Boston	+ 1 617 532 0920
Chicago	+ 1 312 675 0545
Monterrey	+ 52 81 1253 4020
New York	+ 1 212 804 3901
San Francisco	+ 1 415 836 8800
Sao Paulo	+ 55 11 3706 1360
Toronto	+ 1 416 628 1007

EUROPE, MIDDLE EAST & AFRICA

Cape Town	+ 27 21 673 0100
Frankfurt	+ 49 69 133 859 00
Geneva	+ 41 22 817 9777
London	+ 44 20 7618 2222
Milan	+ 39 02 5849 0415
Paris	0800 91 59 17 *

ASIA PACIFIC

China North	10800 852 1032 *
China South	10800 152 1032 *
Hong Kong	+ 852 2844 9333
Mumbai	+ 91 22 6784 9160
Seoul	00798 8521 3392 *
Singapore	800 852 3749 *
Sydney	+ 61 2 9033 9333
Taipei	008 0112 7513 *
Thailand	0018 0015 6207 7181 *
Tokyo	81 3 5290 1555

* = toll free

msci.com

clientservice@msci.com

MSCI

NOTICE AND DISCLAIMER

This document and all of the information contained in it, including without limitation all text, data, graphs, charts (collectively, the "Information") is the property of MSCI Inc. or its subsidiaries (collectively, "MSCI"), or MSCI's licensors, direct or indirect suppliers or any third party involved in making or compiling any Information (collectively, with MSCI, the "Information Providers") and is provided for informational purposes only. The Information may not be modified, reverse-engineered, reproduced or redisseminated in whole or in part without prior written permission from MSCI.

The Information may not be used to create derivative works or to verify or correct other data or information. For example (but without limitation), the Information may not be used to create indexes, databases, risk models, analytics, software, or in connection with the issuing, offering, sponsoring, managing or marketing of any securities, portfolios, financial products or other investment vehicles utilizing or based on, linked to, tracking or otherwise derived from the Information or any other MSCI data, information, products or services.

The user of the Information assumes the entire risk of any use it may make or permit to be made of the Information. NONE OF THE INFORMATION PROVIDERS MAKES ANY EXPRESS OR IMPLIED WARRANTIES OR REPRESENTATIONS WITH RESPECT TO THE INFORMATION (OR THE RESULTS TO BE OBTAINED BY THE USE THEREOF), AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, EACH INFORMATION PROVIDER EXPRESSLY DISCLAIMS ALL IMPLIED WARRANTIES (INCLUDING, WITHOUT LIMITATION, ANY IMPLIED WARRANTIES OF ORIGINALITY, ACCURACY, TIMELINESS, NON-INFRINGEMENT, COMPLETENESS, MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE) WITH RESPECT TO ANY OF THE INFORMATION.

Without limiting any of the foregoing and to the maximum extent permitted by applicable law, in no event shall any Information Provider have any liability regarding any of the Information for any direct, indirect, special, punitive, consequential (including lost profits) or any other damages even if notified of the possibility of such damages. The foregoing shall not exclude or limit any liability that may not by applicable law be excluded or limited, including without limitation (as applicable), any liability for death or personal injury to the extent that such injury results from the negligence or willful default of itself, its servants, agents or sub-contractors.

Information containing any historical information, data or analysis should not be taken as an indication or guarantee of any future performance, analysis, forecast or prediction. Past performance does not guarantee future results.

The Information should not be relied on and is not a substitute for the skill, judgment and experience of the user, its management, employees, advisors and/or clients when making investment and other business decisions. All Information is impersonal and not tailored to the needs of any person, entity or group of persons.

None of the Information constitutes an offer to sell (or a solicitation of an offer to buy), any security, financial product or other investment vehicle or any trading strategy.

It is not possible to invest directly in an index. Exposure to an asset class or trading strategy or other category represented by an index is only available through third party investable instruments (if any) based on that index. MSCI does not issue, sponsor, endorse, market, offer, review or otherwise express any opinion regarding any fund, ETF, derivative or other security, investment, financial product or trading strategy that is based on, linked to or seeks to provide an investment return related to the performance of any MSCI index (collectively, "Index Linked Investments"). MSCI makes no assurance that any Index Linked Investments will accurately track index performance or provide positive investment returns. MSCI Inc. is not an investment adviser or fiduciary and MSCI makes no representation regarding the advisability of investing in any Index Linked Investments.

Index returns do not represent the results of actual trading of investible assets/securities. MSCI maintains and calculates indexes, but does not manage actual assets. Index returns do not reflect payment of any sales charges or fees an investor may pay to purchase the securities underlying the index or Index Linked Investments. The imposition of these fees and charges would cause the performance of an Index Linked Investment to be different than the MSCI index performance.

The Information may contain back tested data. Back-tested performance is not actual performance, but is hypothetical. There are frequently material differences between back tested performance results and actual results subsequently achieved by any investment strategy.

Constituents of MSCI equity indexes are listed companies, which are included in or excluded from the indexes according to the application of the relevant index methodologies. Accordingly, constituents in MSCI equity indexes may include MSCI Inc., clients of MSCI or suppliers to MSCI. Inclusion of a security within an MSCI index is not a recommendation by MSCI to buy, sell, or hold such security, nor is it considered to be investment advice.

Data and information produced by various affiliates of MSCI Inc., including MSCI ESG Research Inc. and Barra LLC, may be used in calculating certain MSCI indexes. More information can be found in the relevant index methodologies on www.msci.com.

MSCI receives compensation in connection with licensing its indexes to third parties. MSCI Inc.'s revenue includes fees based on assets in Index Linked Investments. Information can be found in MSCI Inc.'s company filings on the Investor Relations section of www.msci.com.

MSCI ESG Research Inc. is a Registered Investment Adviser under the Investment Advisers Act of 1940 and a subsidiary of MSCI Inc. Except with respect to any applicable products or services from MSCI ESG Research, neither MSCI nor any of its products or services recommends, endorses, approves or otherwise expresses any opinion regarding any issuer, securities, financial products or instruments or trading strategies and MSCI's products or services are not intended to constitute investment advice or a recommendation to make (or refrain from making) any kind of investment decision and may not be relied on as such. Issuers mentioned or included in any MSCI ESG Research materials may include MSCI Inc., clients of MSCI or suppliers to MSCI, and may also purchase research or other products or services from MSCI ESG Research. MSCI ESG Research materials, including materials utilized in any MSCI ESG Indexes or other products, have not been submitted to, nor received approval from, the United States Securities and Exchange Commission or any other regulatory body.

Any use of or access to products, services or information of MSCI requires a license from MSCI. MSCI, Barra, RiskMetrics, IPD, FEA, InvestorForce, and other MSCI brands and product names are the trademarks, service marks, or registered trademarks of MSCI or its subsidiaries in the United States and other jurisdictions. The Global Industry Classification Standard (GICS) was developed by and is the exclusive property of MSCI and Standard & Poor's. "Global Industry Classification Standard (GICS)" is a service mark of MSCI and Standard & Poor's.

APPENDIX

A non backtest of ES

MSCI

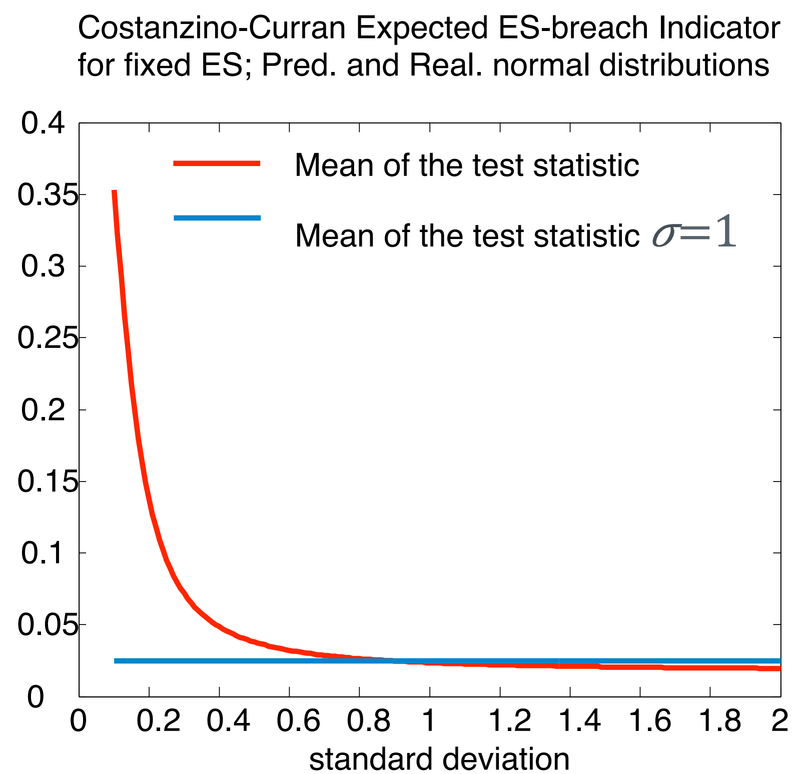
A NON-TEST OF ES

- Costanzino and Curran (2014) and Du and Escanciano (2015) propose an identical "backtest for ES"
- Observing that $ES_{\alpha} = \frac{1}{\alpha} \int_{0}^{\alpha} VaR_{q}\, dq$
- By *analogy*, they define "<u>failure rate of ES</u>" $\frac{1}{\alpha} \int_{0}^{\alpha} (X + VaR_{q} < 0)\, dq$
  - And show it has mean $= \alpha/2$; variance $= \alpha(4-3\alpha)/12$
- Problem:
  - The average of VaR failure rates is not the failure rate of the average VaR
  - The name is deceptive; it doesn't test ES
  - It's a test of the entire tail

MSCI

A NON-TEST OF ES

- Example:
  - $N(0,1)$ predictive distribution for ES_α, $\alpha=5\%$
  - $N(\mu(\sigma),\sigma^2)$, realized distribution
    - $\mu(\sigma)$ such that ES_α coincides with $N(0,1)$
- Prediction of ES_α perfect by construction
- We plot the expected value of the "failure rate of ES" as a function of σ
  - It is different from $\alpha/2=2.5\%$ for all $\sigma\neq1$
  - Shorter tails are penalized by the test and vice versa



MSCI