BACKTESTING EXPECTING SHORTFALL

Introducing three model-independent, non-parametric back-test methodologies for Expected Shortfall

By Carlo Acerbi and Balazs Szekely

NOVEMBER 2014

ABOUT THE AUTHORS

Carlo Acerbi, PhD
EXECUTIVE DIRECTOR, RISK RESEARCH
carlo.acerbi@msci.com

Carlo Acerbi currently works in the MSCI Geneva office as a risk researcher. His main areas of interest in finance are risk management and derivatives pricing.

In the past Dr Acerbi worked as a Risk Manager and as a Financial Engineer for some Italian major banks, and as a senior expert in the risk practice of McKinsey & Co.

Dr Acerbi received a PhD in Theoretical Physics from the International School for Advanced Studies (SISSA - ISAS), Trieste, Italy, before turning to Finance in 1997.

He is the author of several papers in renowned international journals, focusing in particular on the theoretical foundations of financial risk and the extension of portfolio theory to illiquid markets. He has taught "advanced derivatives" at Bocconi University, Milan and is currently an Executive Fellow of the Essex Business School (UK), and a honorary professor of the Corvinus University, Budapest.

He has been for years a member of the board of 'The Journal of Risk'.

Balazs Szekely
SENIOR ASSOCIATE, RISK METHODOLOGY AND LIQUIDITY
balazs.szekely@msci.com

Balazs Szekely currently works in the MSCI Budapest office as a risk researcher. His main area of interest in finance is liquidity risk. He received a PhD in Mathematics in 2005 from Budapest University of Technology and Economics (BUTE). Balazs was a professor in the same institution until 2011, where he conducted research in telecommunications and stochastic processes.

EXECUTIVE SUMMARY

When RiskMetrics announced Value-at-Risk (VaR) as its stated measure of risk in 1996, it initiated an industry standard for institutional risk management which was quickly adopted by the Basel Committee. By 1998, academic researchers began to critique VaR as a risk measure with structural drawbacks, saying it should be replaced by a statistically coherent method. In 2001, an alternative was proposed: Expected Shortfall (ES). Practitioners worldwide adopted ES in parallel with VaR, citing how ES could detect tail risk and offered better mathematical properties for portfolio risk aggregation (i.e., sub-additivity).

For the next 15 years, academic debates over VaR versus ES often reached impassioned levels. During this era, Barra and Riskmetrics both introduced ES in their analytics toolkit, leaving it up to clients to choose between VaR or ES. Until recently, VaR was the measure required by regulators.

However, in October 2013, The Basel Committee on Banking Supervision proposed a fundamental overhaul of its bank trading-book rules after finding discrepancies among banks, in an effort to capture the types of losses lenders might suffer in a period of significant financial stress. Part of this proposal was to change the measurement method for calculating losses from the usual method of VAR to an alternative known as "Expected Shortfall" which regulators believed will better capture the extreme losses that can occur during times of systemic turmoil.

This proposal was criticized though because a research paper published in 2011 had demonstrated that Expected Shortfall (ES) does not possess a mathematical property called elicitability, leading people to believe that that Expected Shortfall could not be back-tested, either.

In this white paper, we join the debate over Expected Shortfall versus VaR by introducing three model-independent, non-parametric back-test methodologies for Expected Shortfall, which we find more powerful than today's standard Basel VaR test. Our three Expected Shortfall back-test's generally require the storage of more information, but we find no conceptual limitations nor computational difficulties. In fact, one of the proposed back tests does not require any additional storage of data from a normal VaR back-test. In addition, we affirm that elicitability has to do with model selection and not model testing, making it almost irrelevant for choosing a regulatory risk standard.

Backtesting Expected Shortfall

Carlo Acerbi[*] *and Balazs Szekely*[†]

MSCI Inc.

October 27, 2014

Abstract

The discovery in 2011 that the Expected Shortfall (ES) is not elicitable, diffused the erroneous belief that it could not be backtested. This misconception aroused a number of criticisms to the recent decision of the Basel Committee to adopt ES in spite of VaR. We contribute to this debate in various ways. First of all, we introduce three model–free, nonparametric backtest methodologies for Expected Shortfall which are shown to be more powerful than the Basel VaR test. These tests generally require the storage of more information, but introduce no conceptual limitations nor computational difficulties of any sort. One of the proposed tests doesn't even require the storage of additional data. Secondly, we observe that elicitability has in fact to do with model *selection* and not with model *testing*, and is therefore irrelevant for the choice of a regulatory risk standard. Finally, we show that ES can in practice be jointly elicited with VaR, but while this may turn out to be a useful result for model selection purposes, we remain convinced that it will not impact the regulatory debate in any respect.

"Eliciwhat?"

Risk professionals had never heard of elicitability until 2011, when [13] proved that Expected Shortfall (ES) is not elicitable as opposed to Value at Risk (VaR). This result sparked a confusing debate.

Put it simply, a statistics $\psi(Y)$ of a random variable Y is said to be elicitable if it minimizes the expected value of a *scoring function* S:

$$\psi = \arg\min_x \mathbb{E}[S(x, Y)]$$

Given a history of point predictions x_t for the statistics and realizations y_t of the random variable, this provides a natural way to evaluate the forecast model, by requiring that the mean score

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

[*]carlo.acerbi@msci.com

[†]balazs.szekely@msci.com

be as low as possible. The mean and the median represent popular examples, minimizing the mean square and absolute error, respectively. The α–quantile, hence VaR, is also elicitable, with score function $S(x,y) = ((x > y) - \alpha)(x - y)$, a well known fact in quantile estimation.

The discovery that ES cannot be elicited led many to conclude that ES would not be backtestable (see for instance [8]) and sounded like the formal proof of a fact that had long been suspected. It is a fact that the absence of a convincing backtest has long been the last obstacle for ES on its way to Basel, as prophesied already in [21].

In October 2013, a consultation paper from the Basel Committee [5] opted to replace VaR with ES for determining the capital charge of internal based models, but kept VaR as the measure to backtest in the usual way. The change was criticized, based on the alleged impossibility to backtest ES, interpreted as a sign that there's something inherently wrong with this risk measure altogether.

Not everyone, however, was convinced. If elicitable means backtestable, how about the few but valuable works on ES backtesting, like [14], which concluded that *"contrary to common belief, ES is not harder to backtest than VaR... Furthermore, the power of the test for ES is considerably higher"*? And what should we do with variance, given that it's not elicitable either? Or why has VaR never been backtested by exploiting its elicitability? At a certain point some dissenting voices started to emerge [11, 20].

In what sense – if any – is it more difficult to backtest ES than VaR? Fundamental reasons? Practical aspects? Power of the test? Model risk? To address these questions we introduce some statistical tests for ES and we compare them with VaR backtests. We restrict our choice to tests which are non–parametric and free from distributional assumptions other than continuity, necessary conditions for any application in banking regulation.

1 Backtesting ES

We adopt a standard hypothesis testing framework for unconditional coverage of ES analogous to the standard Basel VaR setting. We assume that independence of arrival of tail events is tested separately, typically by just visual inspection of VaR exception clusters. This is still the preferred practice in the industry as it provides better insight than proposed tests such as [9, 12].

We assume that every day $t = 1, \ldots, T$, X_t represents a bank's profit–loss distributed along a *real* (unknowable) distribution F_t and it is forecasted by a model *predictive* distribution P_t conditional to previous information used to compute $VaR_{\beta,t}$ and $ES_{\alpha,t}$ as defined by (see [1])

$$ES_{\alpha,t} = -\frac{1}{\alpha}\int_0^\alpha P_t^{-1}(q)dq \tag{1}$$

The random variables $\vec{X} = \{X_t\}$ are assumed to be independent, but not identically distributed. We do not restrict in any respect the variability of F_t and P_t

over time. We will denote by $VaR^F_{\alpha,t}$ and $ES^F_{\alpha,t}$ the value of the risk measures when $X \sim F$.

We assume in what follows that the distributions are continuous and strictly increasing, in which case ES can be expressed as

$$ES_{\alpha,t} = -\mathbb{E}[X_t | X_t + VaR_{\alpha,t} < 0] \tag{2}$$

and VaR is uniquely defined as $VaR_{\beta,t} = -P_t^{-1}(\beta)$. In real cases, this assumption is completely innocuous[1].

Without loss of generality in our numerical examples we will use $T = 250$, $\beta = 1\%$ and $\alpha = 2.5\%$, the relevant case in [5]. $ES_{2.5\%}$ was correctly chosen by the Committee to equal $VaR_{1\%}$ for Gaussian tails, and penalize heavier tails. This is analogous to replacing 50 with 80 in road signs when switching from mph to km/h.

Our null hypothesis generically assumes that the prediction is correct, while the alternative hypotheses are chosen to be only in the direction of risk underestimation. This is again in line with the Basel VaR test, which is meant to detect only excesses of VaR exceptions. We formulate more precise test–specific H_0 and H_1 below. Concrete examples of H_1's will be analyzed in section 2.2 to compute the power of tests in selected cases, similar to the approach followed in table 1 of [5] for different levels of VaR coverage mismatch.

1.1 Test 1: testing ES after VaR

Our first test is inspired by the conditional expectation (2), from which we can easily derive

$$\mathbb{E}\left[\frac{X_t}{ES_{\alpha,t}} + 1 \middle| X_t + VaR_{\alpha,t} < 0 \right] = 0 \tag{3}$$

If $VaR_{\alpha,t}$ has been tested already we can separately test the magnitude of the realized exceptions against the model predictions. Defining $I_t = (X_t + VaR_{\alpha,t} < 0)$, the indicator function of an α–exception, we define the test statistics.

$$Z_1(\vec{X}) = \frac{\sum_{t=1}^{T} \frac{X_t I_t}{ES_{\alpha,t}}}{N_T} + 1 \tag{4}$$

if $N_T = \sum_{t=1}^{T} I_t > 0$.

For this test we choose a null hypothesis

$$H_0 : P_t^{[\alpha]} = F_t^{[\alpha]}, \ \forall t$$

where $P_t^{[\alpha]}(x) = \min(1, P_t(x)/\alpha)$ is the distribution tail for $x < -VaR_{\alpha,t}$. The alternatives are

$$\begin{aligned} H_1 : \quad & ES^F_{\alpha,t} \geq ES_{\alpha,t}, \ \text{for all } t \text{ and } > \text{for some } t \\ & VaR^F_{\alpha,t} = VaR_{\alpha,t}, \text{for all } t \end{aligned}$$

[1]We also assume that $VaR_\alpha > 0$ as it happens in a realistic portfolio p&l distribution

We see that the predicted VaR_α is still correct under H_1, in line with the idea that this test is subordinated to a preliminary VaR test. This test is in fact completely insensitive to an excessive number of exceptions as it's an average taken over exceptions themselves.

Under these conditions $\mathbb{E}_{H_0}[Z_1|N_T > 0] = 0$ and $\mathbb{E}_{H_1}[Z_1|N_T > 0] < 0$ (proposition A.2). So, the realized value $Z_1(\vec{x})$ is expected to be zero, and it signals a problem when it is negative.

Dividing eq. (3) by $ES_{\alpha,t}$ was unnecessary. Normalizing by another statistic of P_t or not normalizing at all, would have given other legitimate tests. Our choice was made to obtain a dimensionless test statistics and to control for heteroscedasticity.

Variations of this test appeared in the literature already several times. For instance, already [18] proposed something similar in a GARCH-EVT context.

1.2 Test 2: testing ES directly

A second test follows from the unconditional expectation

$$ES_{\alpha,t} = -\mathbb{E}\left[\frac{X_t\, I_t}{\alpha}\right] \tag{5}$$

that suggests to define

$$Z_2(\vec{X}) = \sum_{t=1}^{T} \frac{X_t\, I_t}{T\,\alpha\, ES_{\alpha,t}} + 1 \tag{6}$$

Appropriate hypotheses for this test are

$$\begin{aligned} H_0: &\quad P_t^{[\alpha]} = F_t^{[\alpha]}, \ \forall t \\ H_1: &\quad ES_{\alpha,t}^F \geq ES_{\alpha,t}, \ \text{for all } t \text{ and } > \text{for some } t \\ &\quad VaR_{\alpha,t}^F \geq VaR_{\alpha,t}, \text{for all } t \end{aligned}$$

We have again $\mathbb{E}_{H_0}[Z_2] = 0$ and $\mathbb{E}_{H_1}[Z_2] < 0$ (proposition A.3). Remarkably, these results do not require independence of the X_t's. Furthermore, the test can be immediately extended to general, non–continuous distributions, by replacing I_t with

$$I'_t = (X_t + VaR_{\alpha,t} < 0) + \frac{\alpha - Prob[X_t + VaR_{\alpha,t} < 0]}{Prob[X_t + VaR_{\alpha,t} = 0]}(X_t + VaR_{\alpha,t} = 0);$$

see eq. (4.12) in [1].

Test 2 jointly evaluates frequency and magnitude of α–tail events as shown by the relationship

$$Z_2 = 1 - (1 - Z_1)\frac{N_T}{T\,\alpha} \tag{7}$$

remembering that $\mathbb{E}_{H_0}[N_T] = T\,\alpha$.

We remark that both tests 1 and 2 might have been defined under the weaker null hypothesis

$$H'_0: \quad \begin{array}{l} ES^F_{\alpha,t} = ES_{\alpha,t}, \text{ for all } t \\ VaR^F_{\alpha,t} = VaR_{\alpha,t}, \text{for all } t \end{array} \tag{8}$$

all the above results holding true. This choice however would have not been sufficient to simulate the test statistics and compute p-values (see section 2).

1.3 Test 3: estimating ES from realized ranks

Following [10, 7] it is possible to backtest the tails of a model by checking if the observed ranks $U_t = P_t(X_t)$ are i.i.d. $U(0,1)$ as they should if the model distribution is correct. To convert this idea into a specific test for ES, we must assign to each quantile its dollar importance that depends on the shape of the tail itself. To this purpose, denoting with[2]

$$\widehat{ES}^{(N)}_\alpha(\vec{Y}) = -\frac{1}{[N\alpha]}\sum_i^{[N\alpha]} Y_{i:N} \tag{9}$$

an ES estimator based on a vector of N i.i.d draws $\vec{Y} = \{Y_i\}$, we define

$$Z_3(\vec{X}) = -\frac{1}{T}\sum_{t=1}^{T} \frac{\widehat{ES}^{(T)}_\alpha(P_t^{-1}(\vec{U}))}{\mathbb{E}_V\left[\widehat{ES}^{(T)}_\alpha(P_t^{-1}(\vec{V}))\right]} + 1 \tag{10}$$

where $\vec{V}$ are i.i.d. $U(0,1)$. The idea is that the entire vector of ranks $\vec{U} = \{U_t\}$ is reused to estimate ES in every past day t, and the result is then averaged over the entire period[3]. In the denominator we don't have $ES_{\alpha,t}$, but a finite sample estimate to compensate for the bias of estimator (9). The denominator can be computed analytically via

$$\mathbb{E}_V\left[\widehat{ES}^{(T)}_\alpha(P_t^{-1}(\vec{V}))\right] = -\frac{T}{[T\alpha]}\int_0^1 I_{1-p}(T-[T\alpha],[T\alpha])\,P_t^{-1}(p)\,dp \tag{11}$$

where the function $I_x(a,b)$ is a regularized incomplete beta function (proposition A.5).

Also in this case $\mathbb{E}_{H_0}[Z_3] = 0$ and $\mathbb{E}_{H_1}[Z_3] < 0$ (proposition A.4). However, the hypotheses involve this time the entire distributions:

$$\begin{array}{ll} H_0: & P_t = F_t, \ \forall t \\ H_1: & P_t \succeq F_t, \text{ for all } t \text{ and } \succ \text{ for some } t \end{array}$$

where $(\succeq)\ \succ$ denotes (weak) first order stochastic dominance.

Test 3 is less natural than tests 1 and 2, but it is extremely general. A similar test may be designed for any other conceivable statistics for which an estimator is available.

[2] $[x]$ is the integer part of x and $Y_{i:N}$ denotes order statistics.

[3] We could as well have chosen just the distribution $P_{t^\star}$ of a specific day, for instance the last one $t^\star = T$.

2 Test Significance and Power

2.1 Significance

For all tests $Z = Z_i$ we simulate the distribution P_Z under H_0 to compute the p–value $p = P_Z(Z(\vec{x}))$ of a realization $Z(\vec{x})$:

$$\begin{array}{lll} \text{simulate independent} & X_t^i \sim P_t, & \forall t, \forall i = 1, \ldots, M \\ \text{compute} & Z^i = Z(\vec{X}^i) & \\ \text{estimate} & p = \sum_{i=1}^{M}(Z^i < Z(\vec{x}))/M & \end{array} \tag{12}$$

where M is a suitably large number of scenarios. Given a preassigned significance level ϕ, the test is finally accepted or rejected if $p \gtrless \phi$.

From the above procedure, we see that while to backtest VaR exceptions it's sufficient to record a single number I_t per day, to backtest ES it may be necessary to keep memory of the entire predictive distributions P_t.

In reality, for Z_1 and Z_2 it is sufficient to record only the α–tail $P_t^{[\alpha]}$ of the predictive distributions, because $X_t I_t$ can be simulated after $I_t \sim Bernoulli(\alpha)$. We will see in section 2.3 that in fact Z_2 lends itself to implementations that do not even require the recording of the predictive distributions.

Storage of more information (a cumulative distribution function per day) is the only difference between backtesting ES and VaR. A practical difference only which poses no technological challenge.

2.2 Power

In the next subsections we run a number of experiments to evaluate the power of the $ES_{2.5\%}$ tests and compare it to the power of the Basel $VaR_{1\%}$ test under selected hypotheses. The examples are based on Student-t distributions which allow to span all possible fat tails indexes. Figure 1 at page 21 shows how to read the results of every experiment. The green vertical lines in the following plots correspond to exactly 5% and 10% significance level, while the black vertical lines are the corresponding closest discrete levels attainable by the VaR test.

The results are summarized in tables in which the left part describes the setup of H_0 and H_1 and the right part the power of tests. Every row in the tables corresponds to one of the significance levels attainable by the VaR test.

Z_1 is not applicable to the examples in which $VaR_{2.5\%}$ varies across the alternatives.

2.2.1 Scaled distributions: ES coverage

We assume that H_1 is a rescaled version of the H_0 distribution: $F(x) = P(x/\gamma)$, $\gamma > 1$, as shown in figure 2. We assume certain levels of ES coverage mismatch, assuming $ES_\alpha^P = ES_{\alpha'}^F$ for $\alpha' = 5\%, 10\%$ so that $\gamma = ES_\alpha^P / ES_{\alpha'}^P$. The results are shown in table 1 and in figures 3 and 4 in which H_0 is chosen to be $\nu = 100$ and $\nu = 5$ respectively.

In these cases, Z_2 clearly outperforms the VaR test in terms of power. Z_3 has instead slightly less power for smaller tail index.

ν	α' (%)	$VaR_{1\%}$	$ES_{2.5\%}$
5	2.5	3.36	3.52
	5	4.10	4.29
	10	5.14	5.38
100	2.5	2.36	2.37
	5	2.68	2.70
	10	3.16	3.18

				Power		
ν	Significance level (%)	Coverage α' (%)	Scale γ (%)	Z_2 (%)	Z_3 (%)	$VaR_{1\%}$ (%)
5	4.1	5	21.9	51.8	25.2	37.4
		10	53	98.5	78.3	93.5
	10.6	5	21.9	69.0	46.4	55.7
		10	53	99.5	92.9	97.3
100	4.0	5	13.7	47.1	39.0	38.8
		10	34	97.4	94.1	94.2
	10.8	5	13.7	64.7	59.1	56.3
		10	34	99.0	98.1	97.6

Table 1: Power of multiple tests for scaled distributions with different ES coverage as explained in Section 2.2.1.

2.2.2 Student-t distributions

We choose H_1 to be a Student-t distribution with smaller ν than H_0, see figure 5. Notice that in this way the variance will also be larger, as $\sigma^2 = \nu/(\nu-2)$. We analyze $\nu = 100, 10, 5, 3$ in figure 6 and $\nu = 10, 5, 3$ in figure 7. We analyze two H_0's with tail-indexes $\nu = 10, 100$ in table 2.

In this case, Z_3 is the most powerful, followed by Z_2 and by the VaR test.

2.2.3 Normalized Student-t distributions

We repeat the previous experiment using normalized Student-t distributions with unit variance (figure 5). In this case the difference between H_0 and H_1 is only due to tail properties and not larger variance. The results are reported in figures 9 and 10 and in the bottom half of table 2.

This case is particularly subtle. Both Z_2 and the VaR test display very little power at all, with VaR doing slightly better. Z_3, on the contrary, performs quite well.

2.2.4 Fixed $VaR_{2.5\%}$ Student-t distributions

In order to analyze Z_1, we repeat experiments 2.2.2 and 2.2.3 also shifting the H_1 distributions in such a way to leave $VaR_{2.5\%}$ unchanged, as in figure 11. The results are reported in table 3 and in figures 12 and 13.

	ν	$VaR_{1\%}$	$ES_{2.5\%}$
Student-t	100	2.36	2.37
	10	2.76	2.81
	5	3.36	3.52
	3	4.54	5.04
Normalized Student-t	100	2.34	2.35
	10	2.47	2.52
	5	2.60	2.72
	3	2.62	2.90

				Power		
	ν in H_0	Significance level (%)	ν in H_1	Z_2 (%)	Z_3 (%)	$VaR_{1\%}$ (%)
Student-t	10	4.0	5	43.4	48.9	37.7
			3	92.3	94.0	87.1
		10.6	5	61.3	66.1	55.5
			3	96.5	97.1	93.5
	100	4.1	10	40.9	54.8	38.2
			3	99.3	99.8	98.5
		10.4	10	57.7	67.7	56.3
			3	99.6	99.9	99.5
Normalized Student-t	10	4.4	5	7.8	18.7	9.0
			3	8.6	31.4	7.4
		11.2	5	16.5	30.6	18.7
			3	16.0	41.1	16.8
	100	4.4	10	8.2	22.1	10.5
			3	12.3	49.1	12.0
		11.0	10	17.9	34.3	21.6
			3	20.5	56.6	24.5

Table 2: Power of multiple tests for varying tail indexes in H_1. Student-t (Subsection 2.2.2) and normalized Student-t (Subsection 2.2.3) distributions are investigated.

Also in this case, Z_2 and the VaR test display modest power. On the other hand, both Z_1 and Z_3 perform very well.

2.2.5 Comment to results

In these, and other experiments that have been performed, Z_2 has proven the most powerful in the case of alternative hypotheses with different volatility, while Z_3 and Z_1 were the most powerful in the case of different tail index. The VaR exceptions test is generally significantly less powerful.

2.3 Avoiding storage of predictive distributions for Z_2

The critical levels for Z_2 display remarkable stability across different distribution types. Table 4 illustrates the levels for 5% and 0.01% significance, the same as in the Basel traffic–light mechanism, for Student-t distributions with different ν and mean. It is clear that a traffic–light based on Z_2 with fixed levels $Z_2^\star = -0.7$ and $Z_2^\star = -1.8$ would perfectly do in all occasions. Notice that the ± 1 location

	ν	$VaR_{1\%}$	$ES_{2.5\%}$
Student-t	100	2.36	2.37
	10	2.51	2.57
	5	2.77	2.93
	3	3.34	3.84
Normalized Student-t	100	2.34	2.35
	10	2.44	2.49
	5	2.57	2.70
	3	2.74	3.03

				Power			
	ν in H_0	Significance level (%)	ν in H_1	Z_1 (%)	Z_2 (%)	Z_3 (%)	$VaR_{1\%}$ (%)
Student-t	10	4.1	5	28.6	11.1	27.4	12.0
			3	72.7	28.8	62.8	24.9
		10.7	5	43.7	20.4	39.1	24.4
			3	82.2	39.8	70.6	41.6
	100	4.3	10	28.2	7.7	25.1	11.0
			3	91.7	38.5	79.5	33.6
		10.9	10	43.2	15.9	36.3	22.1
			3	94.4	49.1	83.3	50.8
Normalized Student-t	10	4.2	5	20.1	7.9	19.0	8.7
			3	44.7	16.0	39.3	13.8
		11.4	5	33.5	16.8	29.9	18.8
			3	58.5	27.5	50.2	26.9
	100	4.1	10	21.2	6.0	18.9	8.3
			3	70.3	19.6	59.8	20.7
		11.1	10	35.2	13.7	29.4	18.6
			3	79.2	31.4	67.4	35.9

Table 3: Power of multiple tests in the experiment of Section 2.2.4 similar to table 2 but with distributions with fixed $VaR_{2.5\%}$.

shifts span an unrealistically large region for a real profit–loss distribution, which is expected to be centered around zero. Notice also that the thresholds deviate significantly only for dramatically heavy tailed distributions, with $\nu = 3$, and that in this case the proposed test would be more penalizing, which is probably a good thing, given that such tails represent a problem by themselves.

The important fact behind this stability is that for implementing Z_2 there is effectively no need to do a MC test and therefore no need to store predictive distributions. Testing Z_2 requires to record only two numbers per day, the magnitude $X_t I_t$ of a $VaR_{\alpha,t}$ exception and the predicted $ES_{\alpha,t}$.

3 Back to elicitability

Now that we have seen that elicitability is not necessary for backtesting, we argue something more: elicitability has in fact nothing to do with backtesting.

	Significance					
	5%			0.01%		
	location			location		
ν	-1	0	1	-1	0	1
3	-0.78	-0.82	-0.88	-3.9	-4.4	-5.5
5	-0.72	-0.74	-0.78	-1.9	-2.0	-2.3
10	-0.70	-0.71	-0.74	-1.8	-1.9	-1.9
100	-0.70	-0.70	-0.72	-1.8	-1.8	-1.9
Gaussian	-0.70	-0.70	-0.72	-1.8	-1.8	-1.9

Table 4: 5% and 0.01% significance thresholds for Z_2 across Student-t distributions with different ν and location

3.1 Model selection, not model testing

Elicitability allows to compare in a natural way (yet not the only way) different models that forecast a statistics in the exact same sequence of events, while recording only point predictions. For instance, if a bank A has multiple VaR models in place for its p&l, the mean score can be used to select the best in class. But this is *model selection*, not *model testing*. It's a relative ranking not an absolute validation.

Regulators on the contrary need to validate individual models from different banks on an absolute scale. To this purpose, elicitability is of no use. A hypothesis test based on elicitability would still require either the collection of the predictive distributions or strong distributional assumptions, with no guarantee of better power a priori.

It is then no coincidence that despite VaR being elicitable, VaR backtests are still based on counting exceptions. If these tests are simple and entail the recording of just one number, it's not because VaR is elicitabile, but because quantiles define a Bernoulli random variable. Any other elicitable statistic simply does not.

3.2 Expectiles

Expectiles [19] have recently attracted a lot of interest (see for instance [17]) because they are the only coherent law–invariant measure of risk which is also elicitable [6, 22]. But if, as we have seen, absence of elicitability is not a serious problem for a regulatory risk standard, absence of comonotonic additivity certainly is. An expectile ρ will tell you that a long position in a call option C is partially hedged by a long (yes, long) position in the underlying stock S: $\rho(C+S) < \rho(C) + \rho(S)$. Even VaR would look perplex.

The class of comonotonic additive coherent measures of risk of law–invariant type has been completely classified [15] and coincide with spectral measures of risk [2, 4], which contain ES as the most popular example. Alternative choices that are not law–invariant belong to the realm of stress–test based measures, which is incidentally the direction the FED seems to be considering with increasing interest.

3.3 Joint elicitability of ES and VaR

An intuitive if not rigorous way to understand why ES is not elicitable is to notice that there exists no expression of the type

$$\mathbb{E}[L(X, ES)] = 0$$

where L is a function involving only a random variable and its ES. If such a function existed we could interpret it as

$$L(X, ES) = \left.\frac{\partial S(X, e)}{\partial e}\right|_{e=ES}$$

and integrate it with respect to e to build a scoring function S that elicits ES. However, there exist null expectations that involve both ES and VaR. For instance

$$\begin{aligned} &\mathbb{E}[(X + ES)(X + VaR < 0)] = 0 \\ &\mathbb{E}[X(X + VaR < 0) + \alpha ES] = 0 \end{aligned} \tag{13}$$

It is therefore clear that if there's a chance to build a scoring function for ES, this needs to involve VaR as well. Starting from the above expressions, it is in fact not difficult to construct a one–parameter family of scoring functions

$$S^W(v, e, x) = \alpha e^2/2 + W\alpha v^2/2 - \alpha ev + \big(e(v + x) + W(x^2 - v^2)/2\big)\,(x + v < 0) \tag{14}$$

for every $W \in \mathbb{R}$, that jointly elicit VaR and ES

$$\{VaR, ES\} = \arg\min_{v,e} \mathbb{E}[S^W(v, e, X)] \tag{15}$$

under the condition that $VaR \cdot W > ES$. Notice that for any fixed W one can imagine a bizarre distribution (e.g. a $\nu = 1 + \epsilon$ Student–t with $\epsilon > 0$ small enough) that violates this condition, so strictly speaking, this is not a mathematical proof of 2–elicitability in the sense of [16], as was proven for variance and mean. However, from a practical point of view, it is easy to choose a value for W large enough for any specific case at hand.

As a side remark, we observe that theoretical results showing that a measure is not elicitable may still not preclude that it is *in practice* elicitable. We still don't know whether VaR and ES are jointly elicitable or not, and we wouldn't be surprised to discover that they are not, but we already know that in practice they are.

We note that this result opens new ways to set up selections for ES models, but in the light of the observations made in 3.1 it does not add anything to ES as a candidate for regulatory standards.

4 Conclusions

ES can be backtested. The most important contribution of our work is to define three ES backtest methods which are non–parametric, distribution independent, and do not assume any asymptotic convergence. The tests are easy to implement and generally display better power than the standard Basel VaR backtest. The only additional complexity they bring about is the necessity to record the predicted cumulative distribution function day by day. This is even unnecessary for Z_2, which exhibits a remarkable stability of the critical levels across different tail shapes.

Elicitability of a risk measure is not relevant for an absolute model validation. This property is useful for relative comparison of different models forecasting the same process, namely for model selection. Non elicitability of a risk measure does not preclude the possibility to backtest it efficiently and elicitability of VaR will never provide a better alternative to backtest it by just counting exceptions.

We provide some intuition as to why ES is not individually elicitable. From this fact, we learn how to build a scoring functional that jointly elicits ES and VaR. The result is new and generally important for ES model selection, but we don't think it will impact, in any respect, the regulatory debate around VaR and ES.

We believe that Z_1, together with the standard Basel VaR backtest or alternatively Z_2 alone, represent valid proposals to backtest models for ES based regulation. Z_3 is also a valid test, but it seems more appropriate as a complementary test to detect tail index mispecification only.

Acknowledgments: We are indebted to Tilmann Gneiting, Dirk Tasche and an anonymous referee for many corrections and enriching discussions. We are grateful to Imre Kondor for spurring this research and for organizing the "International workshop on systemic risk and regulatory market risk measures" in Pullach, where a preliminary version of this work was first presented.

References

[1] Acerbi, C. and Tasche, D. (2002) On the coherence of expected shortfall, Journal of Banking & Finance **26** (1487–1503)

[2] Acerbi, C. (2002) Spectral measures of risk: A coherent representation of subjective risk aversion, Journal of Banking & Finance **26** (1505–1514)

[3] Acerbi, C. (2004) Coherent representations of subjective risk aversion, in Risk Measures for the 21^{st} Century, Wiley

[4] Acerbi, C. (2007) Coherent measures of risk in everyday market practice, Quantitative Finance, **7**, 4, (359-364)

[5] Basel Committee on Banking Supervision (2013) Fundamental review of the trading book: A revised market risk framework, Second Consultative Paper.

[6] Bellini, F., Klar, B., Müller, A. and Rosazza Gianin, E. (2014) Generalized quantiles as risk measures. Insurance: Mathematics and Economics **54**, (41–48)

[7] Berkowitz, J., (2001) Testing density forecasts, applications to risk management. J. Bus. Econ. Stat. **19**, (465–474).

[8] Carver. L. (2013) Mooted VAR substitute cannot be back-tested, says top quant, RISK, March 2013

[9] Christoffersen, P., (1998) Evaluating interval forecasts. Int. Econ. Rev. **39**, (841–862).

[10] Diebold, F.X., Gunther, T., Tay, A., (1998). Evaluating density forecasts, with applications to financial risk management. Int. Econ. Rev. **39**, (863–883).

[11] Emmer, S., Kratz, M. and Tasche, D. (2013) What is the best risk measure in practice? A comparison of standard measures, working paper.

[12] Engle, R., Manganelli, S. (2004). CAViaR: Conditional value at risk by quantile regression. J. Bus. Econ. Stat. **22**, (367–381).

[13] Gneiting, T. (2011) Making and Evaluating Point Forecasts, Journal of the American Statistical Association **106**, (494).

[14] Kerkhof, J. and Melenberg, B. (2004) Backtesting for Risk-Based Regulatory Capital, J. Bank. Finance **28** (1845–1865).

[15] Kusuoka, S. (2001) On law invariant coherent risk measures. Adv. Math. Econ. **3**, 83-95

[16] Lambert, N., Pennock, D. M., Shoham, Y. (2008) Eliciting Properties of Probability Distributions, Proceedings of the 9^{th} ACM Conference on Electronic Commerce, EC 08

[17] Martin, R. (2014) Expectiles behave as expected, Risk Magazine **27** 6, (79-83)

[18] McNeil, A. J., Frey, R. (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, Journal of Empirical Finance **7** (217-300)

[19] Newey, W. K. and Powell, J. L. (1987) Asymmetric least squares estimation and testing. Econometrica, **55** (819-847)

[20] Tasche, D. (2013) Expected Shortfall is not elicitable. So what? Presentation at Imperial College, London.

[21] Yamai, Y. and Yoshiba, T. (2002) On the Validity of Value-at-Risk: Comparative Analyses with Expected Shortfall, Monetary and economic studies **20.1**, (57–85).

[22] Ziegel, J. F. Coherence and elicitability, to appear on Mathematical Finance, preprint available at `http://arxiv.org/abs/1303.1690v2`

A Proofs

We provide the proof to some key results used in the article. First of all we prove the following general

Lemma A.1. *Let X_i, $i = 1, \ldots, T$ be i.i.d. draws from a continuous and strictly increasing distribution F and assume $VaR_\alpha = -F^{-1}(\alpha)$ is known. Denote as usual $I_i = (X_i + VaR_\alpha < 0)$ and $N_T = \sum_{i=1}^T I_i$. Define the estimator*

$$\widehat{ES}_\alpha^{(N)}(X) = -\frac{\sum_{i=1}^N X_i\, I_i}{N_T} \tag{16}$$

if $N_T > 0$.

Then, $\widehat{ES}_\alpha^{(N)}(X)$ is a conditionally unbiased estimator of $ES_\alpha(X)$, that is,

$$\mathbb{E}\left[\widehat{ES}_\alpha^{(N)}(X) \middle| N_T > 0\right] = ES_\alpha. \tag{17}$$

Proof: Conditioning first on the I_i's and then using the independence of the X_i, we get

$$\begin{aligned}
\mathbb{E}\left[\widehat{ES}_\alpha^{(N)} \middle| N_T > 0\right] &= \mathbb{E}\left[-\frac{1}{N_T}\sum_i X_i\, I_i \middle| N_T > 0\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[-\frac{1}{N_T}\sum_i X_i\, I_i \middle| I_1, \ldots, I_T\right] \middle| N_T > 0\right] \\
&= \mathbb{E}\left[-\frac{1}{N_T}\sum_i I_i\, \mathbb{E}[X_i \mid I_i] \middle| N_T > 0\right] \\
&= \mathbb{E}\left[\frac{1}{N_T}\sum_i I_i\, ES_\alpha \middle| N_T > 0\right] \\
&= ES_\alpha.
\end{aligned}$$

Note that $I_i\mathbb{E}[X_i|I_i] = -I_i ES_\alpha$ holds regardless of the value I_i. □

We now consider the setting adopted throughout the paper. We denote independent r.v.'s $X_t \sim F_t$ for $t = 1, \ldots, T$. We denote by $ES_{\alpha,t}$ and $VaR_{\alpha,t}$ the statistics computed from predictive distributions P_t, $I_t = (X_t + VaR_{\alpha,t} < 0)$ the indicator function of a VaR exception and $N_T = \sum_t I_t$ the overall number of exceptions. All distribution functions are assumed to be continuous and strictly increasing. We also assume $VaR_{\alpha,t} > 0$ and a fortiori $ES_{\alpha,t} > 0$ throughout.

The properties of Z_1 are proven below under null hypothesis (8) which is weaker than in section 1.1.

Proposition A.2. *In the test hypotheses*

$$\begin{array}{ll} H_0: & ES^F_{\alpha,t} = ES_{\alpha,t}, \text{ for all } t \\ & VaR^F_{\alpha,t} = VaR_{\alpha,t}, \text{for all } t \\ H_1: & ES^F_{\alpha,t} \geq ES_{\alpha,t}, \text{ for all } t \text{ and } > \text{for some } t \\ & VaR^F_{\alpha,t} = VaR_{\alpha,t}, \text{for all } t \end{array}$$

we have

1. $\mathbb{E}_{H_0}[Z_1|N_T > 0] = 0$
2. $\mathbb{E}_{H_1}[Z_1|N_T > 0] < 0$

Proof: *1.* Similar to lemma A.1. Assuming H_0, implies that $ES_{\alpha,t} = ES^P_{\alpha,t} = ES^F_{\alpha,t}$ and $I_t = I^P_t = I^F_t$

$$\begin{aligned} \mathbb{E}_{H_0}\left[Z_1 | N_T > 0\right] &= \mathbb{E}_{H_0}\left[\frac{1}{N_T}\sum_t \frac{X_t\, I_t}{ES_{\alpha,t}} + 1 \middle| N_T > 0\right] \\ &= \mathbb{E}_{H_0}\left[\mathbb{E}_{H_0}\left[\frac{1}{N_T}\sum_t \frac{X_t\, I_t}{ES_{\alpha,t}} \middle| I_1, \ldots, I_T\right] + 1 \middle| N_T > 0\right] \\ &= \mathbb{E}_{H_0}\left[\frac{1}{N_T}\sum_t I_t\,\frac{\mathbb{E}_{H_0}[X_t \mid I^F_t]}{ES_{\alpha,t}} + 1 \middle| N_T > 0\right] \\ &= \mathbb{E}_{H_0}\left[-\frac{1}{N_T}\sum_t \frac{I_t\, ES^F_{\alpha,t}}{ES_{\alpha,t}} + 1 \middle| N_T > 0\right] \\ &= 0 \end{aligned}$$

where we have used independence of X_t's to condition on a single I_t only.

2. Let's assume H_1. Along the same lines as before, we have

$$\begin{aligned} \mathbb{E}_{H_1}\left[Z_1 | N_T > 0\right] &= \mathbb{E}_{H_1}\left[\frac{1}{N_T}\sum_t I_t\,\frac{\mathbb{E}_{H_1}[X_t \mid I_t]}{ES_{\alpha,t}} + 1 \middle| N_T > 0\right] \\ &= \mathbb{E}_{H_1}\left[\frac{1}{N_T}\sum_t I_t\,\frac{\mathbb{E}_{H_1}[X_t \mid I^F_t]}{ES_{\alpha,t}} + 1 \middle| N_T > 0\right] \\ &= \mathbb{E}_{H_1}\left[-\frac{1}{N_T}\sum_t \frac{I_t\, ES^F_{\alpha,t}}{ES^P_{\alpha,t}} + 1 \middle| N_T > 0\right] \\ &< 0 \end{aligned}$$

Notice that we used the fact that under H_1 we assume $VaR^F_{\alpha,t} = VaR^P_{\alpha,t}$ so that $I^F_t = I^P_t = I_t$. The last step follows from $ES^F_{\alpha,t} \geq ES^P_{\alpha,t} > 0$ which holds for all t and the inequality is strict for some t. □

The properties of Z_2 are proven below under weaker H_0 than in section 1.2.

Proposition A.3. *In the test hypotheses*

$$\begin{array}{ll} H_0: & ES^F_{\alpha,t} = ES_{\alpha,t},\ \textit{for all } t \\ & VaR^F_{\alpha,t} = VaR_{\alpha,t}, \textit{for all } t \\ H_1: & ES^F_{\alpha,t} \geq ES_{\alpha,t},\ \textit{for all } t \textit{ and } > \textit{for some } t \\ & VaR^F_{\alpha,t} \geq VaR_{\alpha,t}, \textit{for all } t \end{array}$$

we have

1. $\mathbb{E}_{H_0}[Z_2] = 0$
2. $\mathbb{E}_{H_1}[Z_2] < 0$

Proof: *1.* From identity (5) we have

$$\mathbb{E}_{H_0}\left[\frac{X_t\, I_t}{\alpha\, ES_{\alpha,t}} + 1\right] = 0, \quad \forall t$$

The conclusion follows from definition (6).

2. The assumption $VaR^P_{\alpha,t} \leq VaR^F_{\alpha,t}$ implies that $I^F_t \leq I_t$. Using this fact and $X_t < 0$ if $I_t = 1$ (here we use $VaR^P_{\alpha,t} > 0$), we get $X_t I_t \leq X_t I^F_t$. Taking expectation, we have

$$\mathbb{E}_{H_1}\left[X_t I_t\right] \leq \mathbb{E}_{H_1}\left[X_t I^F_t\right] = -\alpha ES^F_t \leq -\alpha ES^P_t.$$

Under H_1 the last inequality holds for all t and it is strict for some t. The conclusion then follows from definition (6). □

Remarkably, the previous proposition holds even if the X_t are not independent.

The properties of Z_3 follow from

Proposition A.4. *In the test hypotheses of section 1.3*

1. $\mathbb{E}_{H_0}[Z_3] = 0$
2. $\mathbb{E}_{H_1}[Z_3] < 0$

Proof: *1.* Under H_0 we have $U_t = P_t(X_t) \sim U(0,1)$. The conclusion directly follows from the definition (10).

2. Under H_1, for any fixed t the variables $P_t^{-1}(U) \sim F_t$ are stochastically dominated by $P_t^{-1}(V) \sim P_t$ where $V \sim U(0,1)$. The estimator $\widehat{ES}$ is consistent with stochastic dominance, so $\mathbb{E}_{H_1}[\widehat{ES}^{(T)}_\alpha(P_t^{-1}(\vec{U}))] \geq \mathbb{E}_V[\widehat{ES}^{(T)}_\alpha(P_t^{-1}(\vec{V}))]$ for all t and the inequality is strict for some t. Conclusion then follows directly from definition (10). □

We prove the following result on the bias of the ES estimator

Proposition A.5. *Let Y_i, $i = 1, \ldots, n$ be i.i.d. draws from a continuous and strictly increasing distribution F. The expected value of the estimator (9) can be expressed as*

$$\mathbb{E}\left[\widehat{ES}_\alpha^{(n)}(\vec{Y})\right] = -\frac{n}{[n\alpha]} \int_0^1 I_{1-p}(n - [n\alpha], [n\alpha])\, F^{-1}(p)\, dp \tag{18}$$

where the function $I_x(a, b)$ is a regularized incomplete beta function.

Proof: The distribution density of the order statistics $Y_{k:n}$ is given by

$$f_k(y) = n\, f(y) \binom{n-1}{k-1} F(y)^{k-1}(1 - F(y))^{n-k}$$

where f is the density of F. Some algebra yields

$$\begin{aligned}
\mathbb{E}\left[\widehat{ES}_\alpha^{(n)}(\vec{Y})\right] &= -\frac{1}{[n\alpha]} \sum_{k=1}^{[n\alpha]} \mathbb{E}\left[Y_{k:n}\right] \\
&= -\frac{1}{[n\alpha]} \sum_{k=1}^{[n\alpha]} \int y\, f_k(y)\, dy \\
&= -\frac{n}{[n\alpha]} \sum_{h=0}^{[n\alpha]-1} \binom{n-1}{h} \int y\, F(y)^h (1 - F(y))^{n-h-1}\, f(y)\, dy \\
&= -\frac{n}{[n\alpha]} \int F_B([n\alpha] - 1; n - 1, F(y)) y\, f(y) dy
\end{aligned}$$

where $F_B(k; n; p)$ is the cumulative distribution function of a standard binomial distribution. The conclusion follows from the representation $F_B(k; n; p) = I_{1-p}(n - k; k + 1)$ and a standard change of integration $f(y)dy = dp$. □

Finally, we prove eq. (15) on joint elicitability of ES and VaR

Proposition A.6. *The scoring function (14) jointly elicits VaR and ES in the sense of equation (15) under the condition $W \cdot VaR > ES$*

Proof: Follows from direct computation. Denoting $\star \equiv \{v = VaR, e = ES\}$ one obtains

$$\left.\frac{\partial}{\partial v}\right|_\star \mathbb{E}[S^W(v, e, X)] = \mathbb{E}[(Wv - e)(\alpha - (X + v < 0))]|_\star = 0$$

$$\left.\frac{\partial}{\partial e}\right|_\star \mathbb{E}[S^W(v, e, X)] = \mathbb{E}[\alpha(e - v) + (v + X)(X + v < 0)]|_\star = 0$$

and

$$\left.\frac{\partial^2}{\partial v^2}\right|_\star \mathbb{E}[S^W(v,e,X)] = \mathbb{E}[(W\,v-e)\delta(X+v)]|_\star = (W\,VaR - ES)f(-VaR)$$

$$\left.\frac{\partial^2}{\partial e^2}\right|_\star \mathbb{E}[S^W(v,e,X)] = \alpha$$

$$\left.\frac{\partial^2}{\partial v \partial e}\right|_\star \mathbb{E}[S^W(v,e,X)] = 0$$

which also shows that the stationary point is a saddle when the condition is violated. □

B Figures

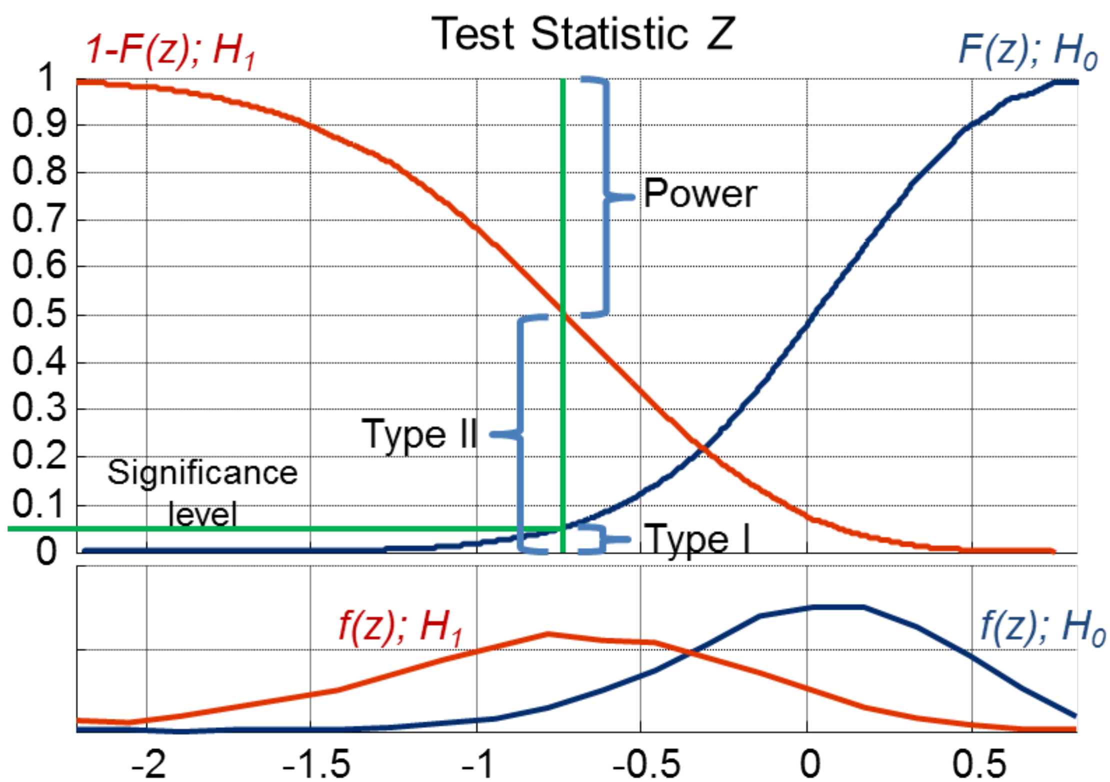

Figure 1: Power and significance for a test under specific H_0 and H_1's. The top plot shows $F(z)$ under H_0 and $1 - F(z)$ under H_1 for the test variable Z. Any vertical line intercepts the two curves at type I probability (the significance level) and type II probability, and displays the test power. The bottom plot shows the densities under H_0 and H_1; more powerful tests result in more distinct curves.

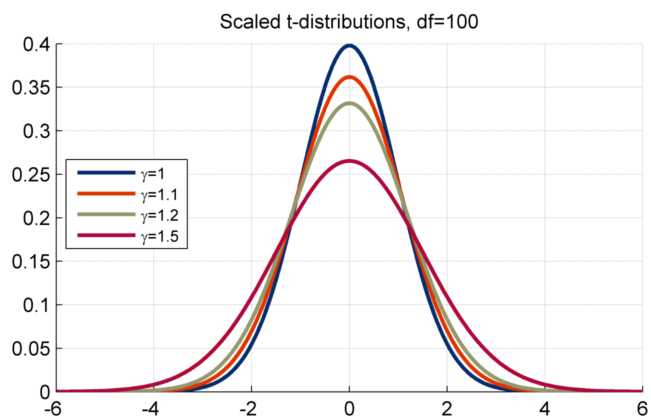

Figure 2: Scaled distributions

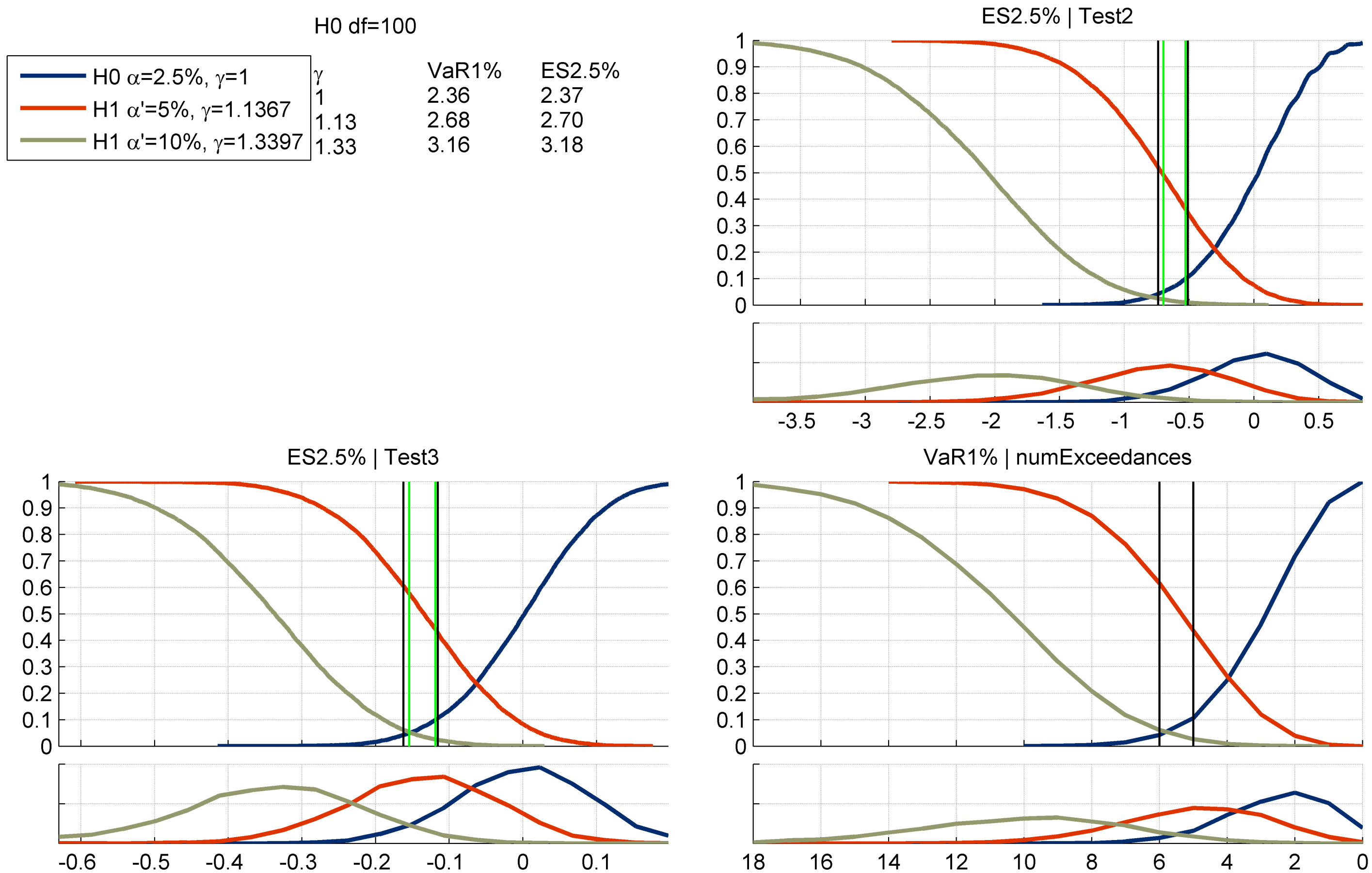

Figure 3: Different ES coverage. H_0: Student–t with $\nu = 100$

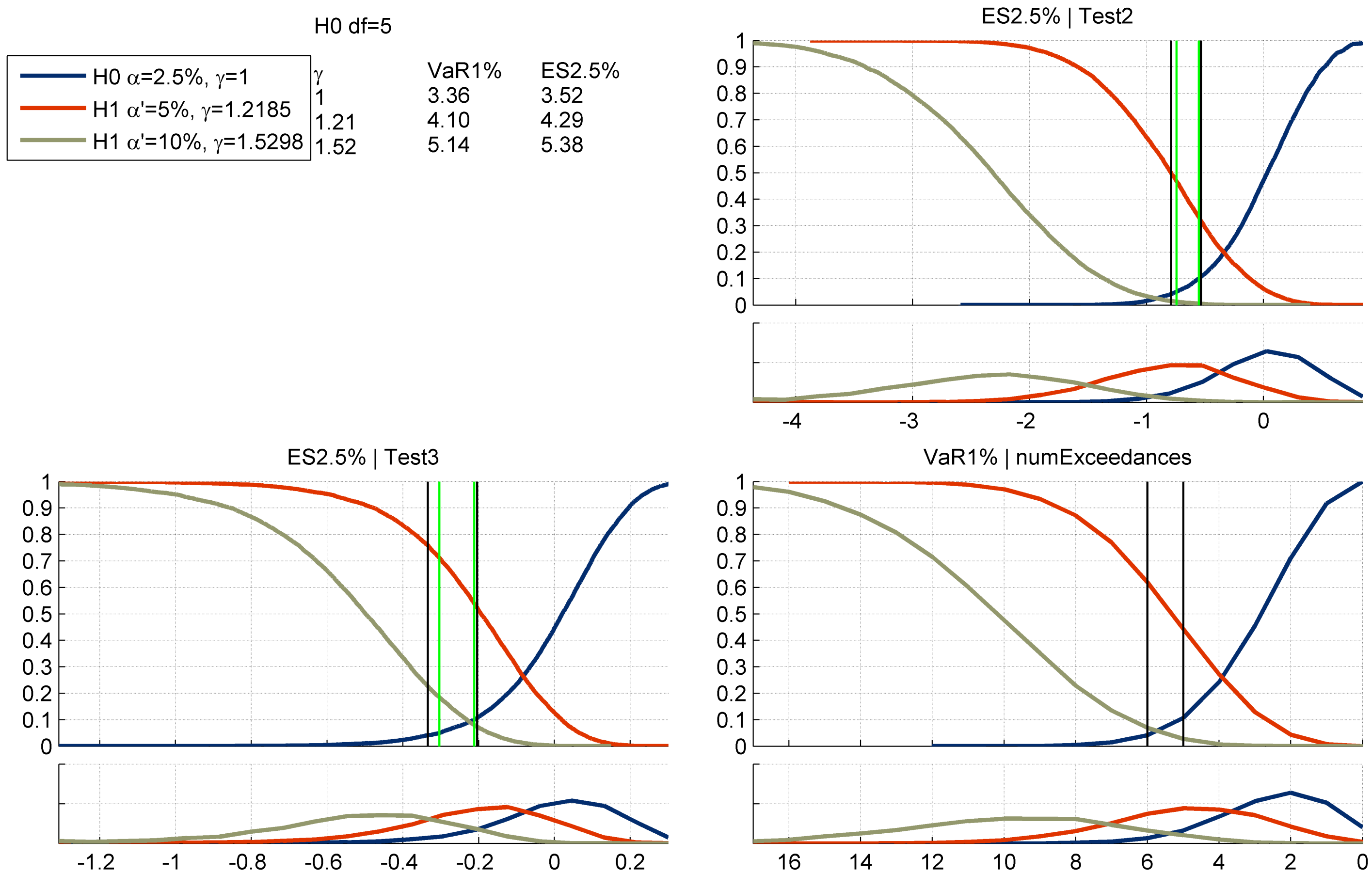

Figure 4: Different ES coverage. H_0: Student–t with $\nu = 5$

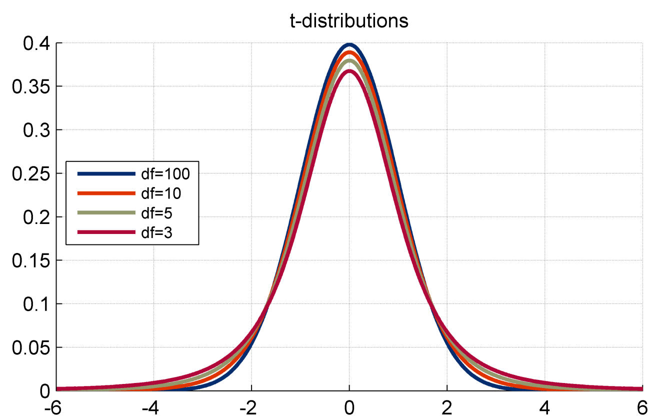

Figure 5: Student-t distributions

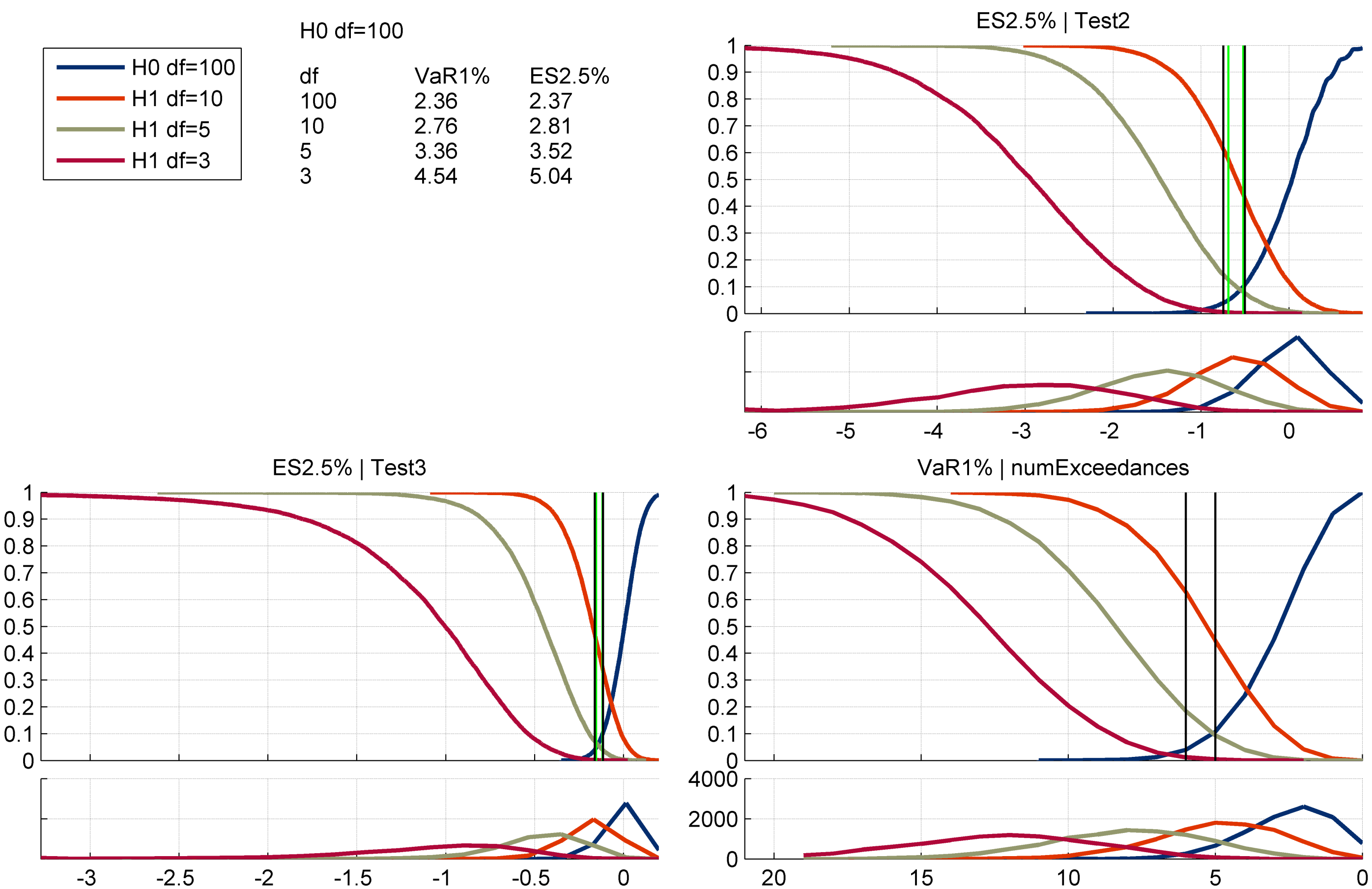

Figure 6: Student–t distributions. H_0: $\nu = 100$. H_1: $\nu = 10, 5, 3$

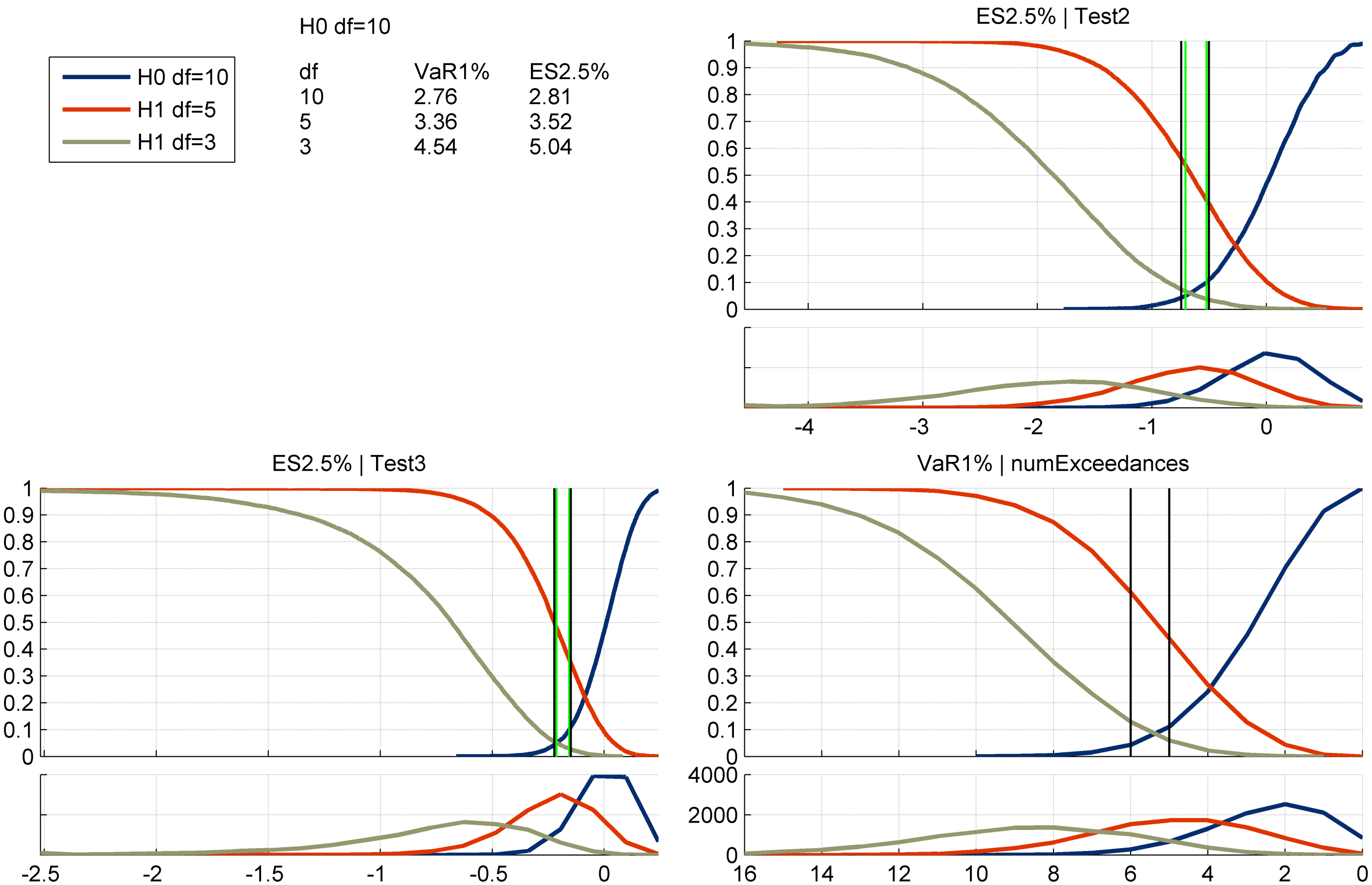

Figure 7: Student–t distributions. H_0: $\nu = 10$. H_1: $\nu = 5, 3$

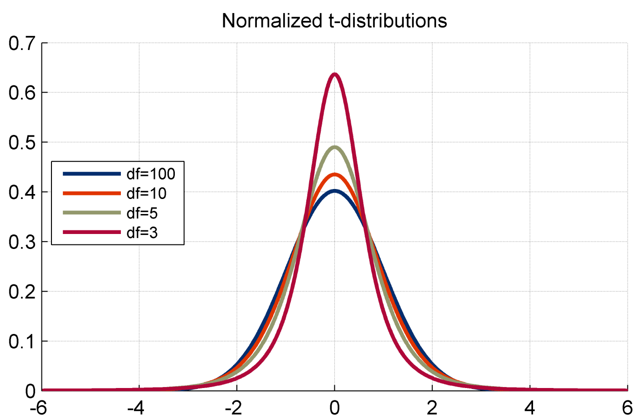

Figure 8: Normalized Student-t distributions

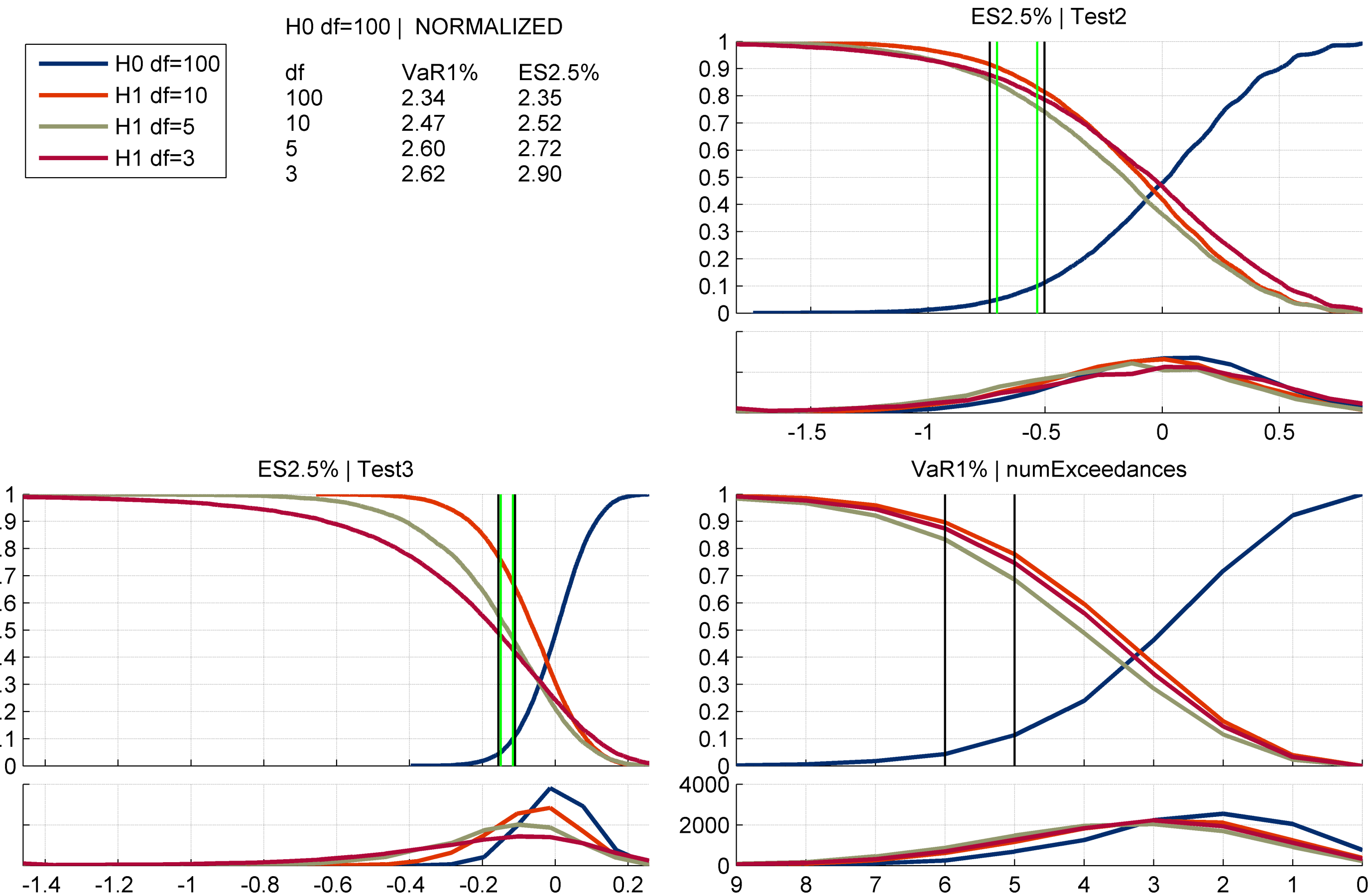

Figure 9: Normalized Student–t distributions. H_0: $\nu = 100$. H_1: $\nu = 10, 5, 3$

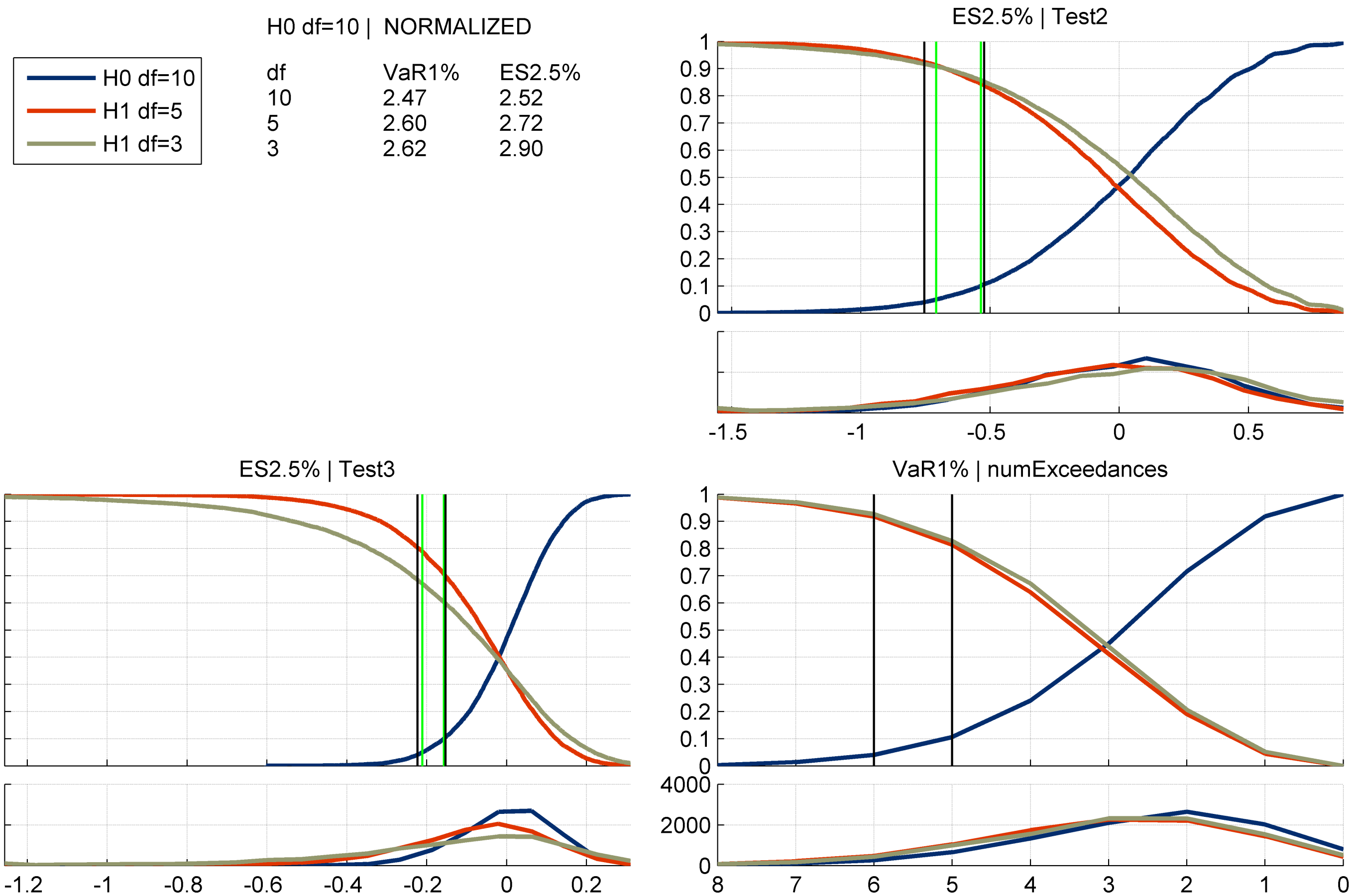

Figure 10: Normalized Student–t distributions. H_0: $\nu = 10$. H_1: $\nu = 5, 3$

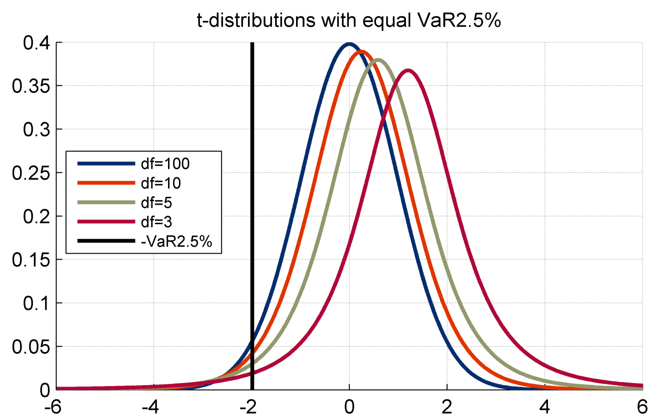

Figure 11: Fixed $VaR_{2.5\%}$ Student-t distributions

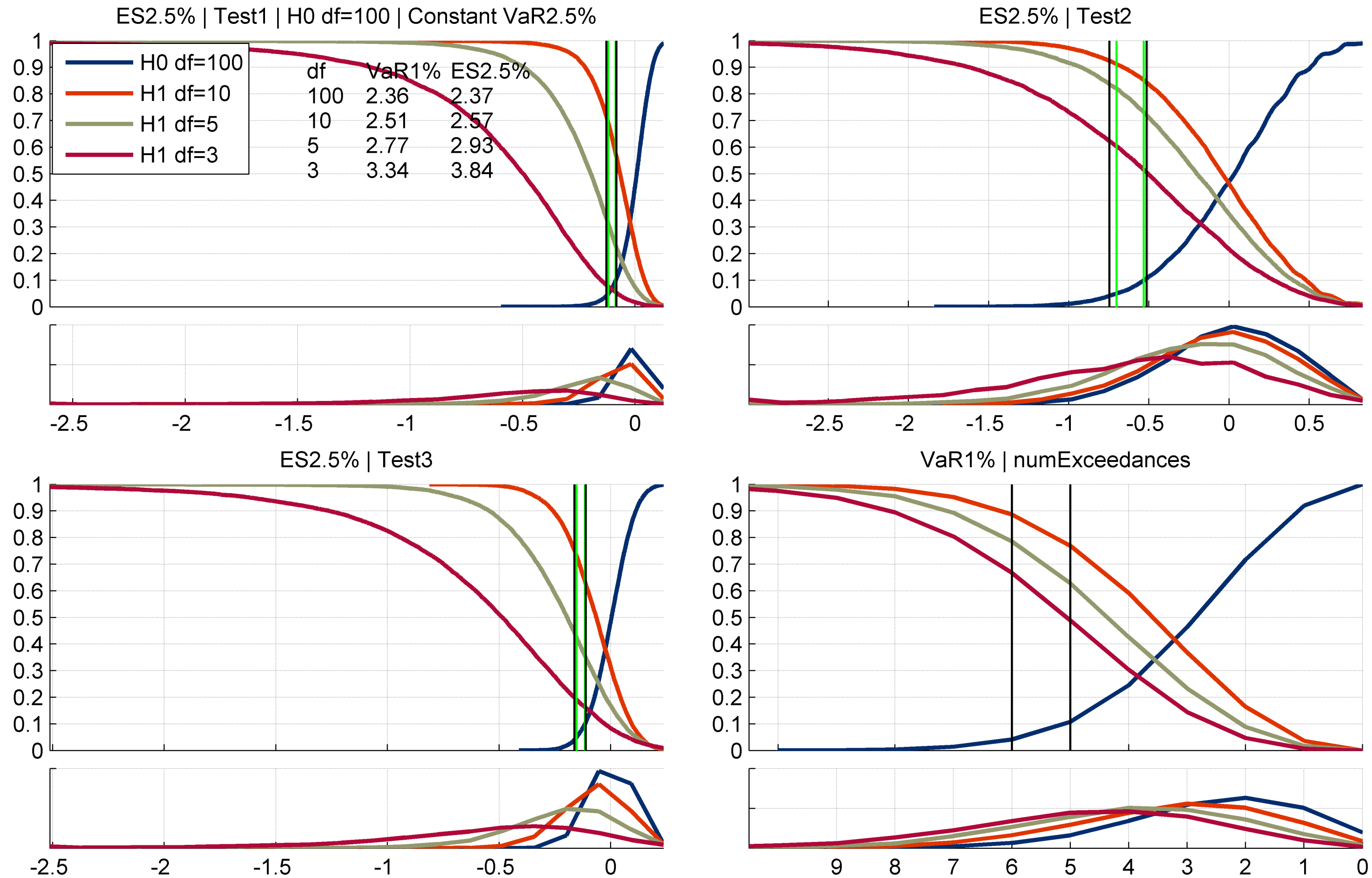

Figure 12: Fixed $VaR_{2.5\%}$ Student–t distributions. H_0: $\nu = 100$. H_1: $\nu = 10, 5, 3$

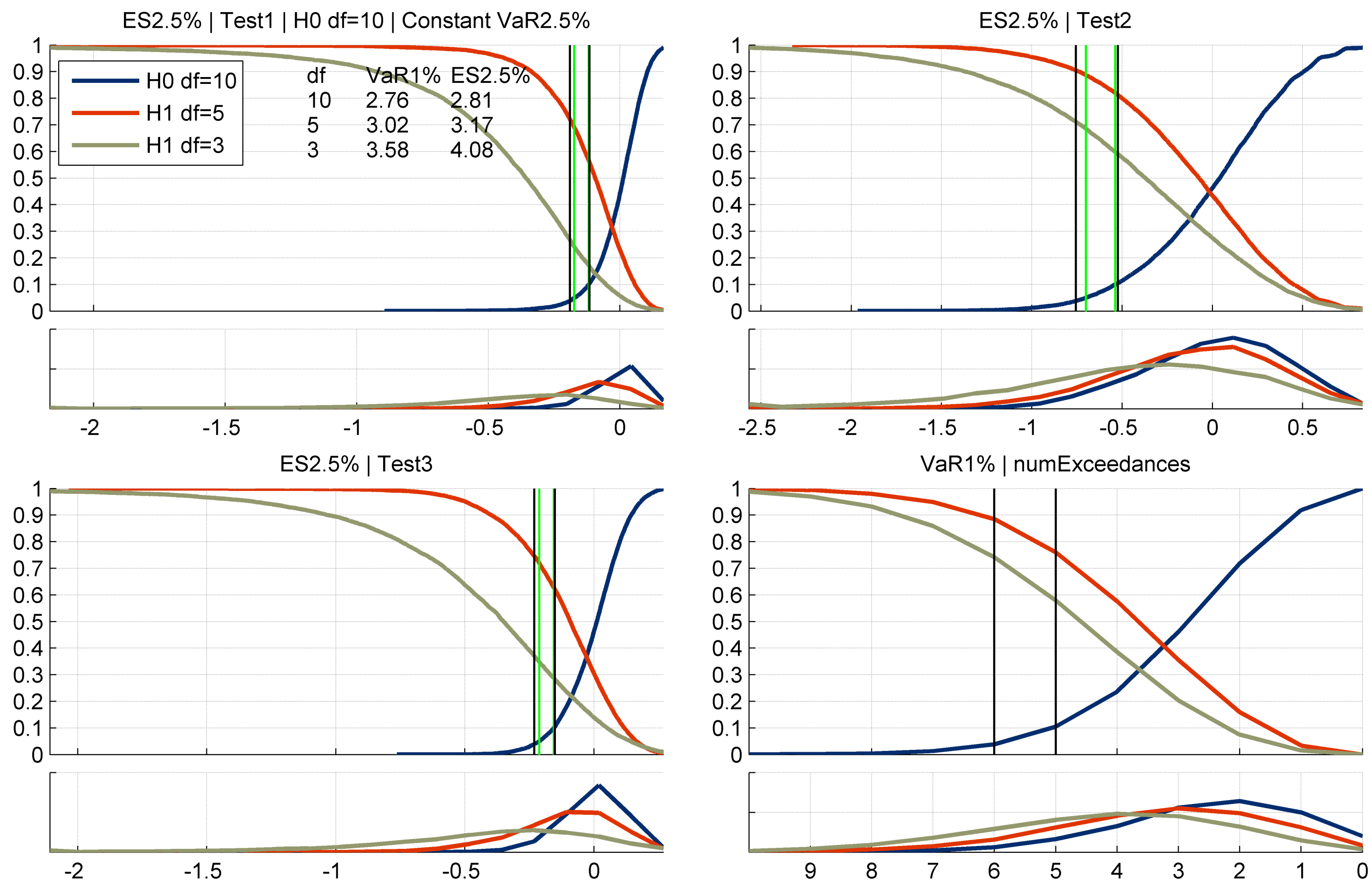

Figure 13: Fixed $VaR_{2.5\%}$ Student–t distributions. H_0: $\nu = 10$. H_1: $\nu = 5, 3$

QUESTIONS ABOUT RISK? CONTACT US

AMERICAS

Americas	1.888.588.4567 (toll free)
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
São Paulo	+ 55.11.3706.1360
Toronto	+ 1.416.628.1007

EUROPE, MIDDLE EAST, AFRICA & INDIA

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
Mumbai	+91.22.6784.9160
Paris	0800.91.59.17 (toll free)

ASIA PACIFIC

China North	10800.852.1032 (toll free)
China South	10800.152.1032 (toll free)
Hong Kong	+ 852.2844.9333
Seoul	00798.8521.3392 (toll free)
Singapore	800.852.3749 (toll free)
Sydney	+ 61.2.9033.9333
Taipei	008.0112.7513 (toll free)
Tokyo	+ 81.3.5290.1555

TO FIND OUT MORE, PLEASE VISIT:
msci.com

CONTACT:
clientservice@msci.com

20 YEARS OF RiSK

RiskMetrics

95 of the top 100 investment managers in the world are MSCI clients.*

* Based on P&I AUM data as of December 2013 and internal MSCI data as of June 2014.

75 of the top 100 global asset managers are MSCI clients.**

** Based on aiCIO data as of Dec 2013 and MSCI data as of June 2014.

msci.com

MSCI