A Breakthrough Idea in Risk Measure Validation – Is the Way Paved for an Effective Expected Shortfall Backtest?*

Gyöngyi Bugár

This research note is a kind of "call for attention" to recent developments in backtesting financial risk measures. This topic is relevant in relation to the regulatory monitoring of the performance of internal risk models used by banks in determining the minimum capital requirements for trading book portfolios. Backtesting is a process for checking the validity of risk estimation models. In his seminal work, Gneiting (2011) has proven that a prominent risk measure, Expected Shortfall (ES), lacks a property called elicitability. This finding has triggered a huge controversy on the issue of whether ES is backtestable at all. Due to the significant contribution of Acerbi and Székely (2017, 2019) among others, the above-mentioned debate can be adequately and convincingly closed because there is a (re)solution. In particular, one can arrive at the conclusion that, building on its joint elicitability with Value-at-Risk (VaR), it is possible to introduce a so-called ridge backtest for ES. In fact, there is still an open question as to when and how the regulatory authorities will (re)act.

1. Introduction

The recently published, new Basel III standards (*BCBS 2016*) provide a revised framework for determining the capital charge for market risk in internal models with a shift from VaR to ES (*Bugár – Ratting 2016*). This is a risk measure for better capturing tail risk which has some more favourable properties such as coherence (see *Artzner et al. 1999*).

However, as *Gneiting* (*2011*) demonstrated, ES has a serious disadvantage compared to VaR because it is not elicitable. This has led some to conclude that this risk measure is not backtestable at all (see for instance, *Carver 2013*). Others such as *Tasche* (*2014*) and *Acerbi – Székely* (*2014*) were definitely convinced that this is not a real problem. The latter authors emphasised that elicitability has not even been exploited in the backtests of VaR. The lack of elicitability in the case of ES has proven to be a challenge and has exploded into a lively debate in the literature, especially between financial mathematicians and statisticians focused on the question of whether ES is a proper risk measure to substitute VaR in banking regulation. One theoretically prominent applicant for this purpose might be the α-expectile which is both elicitable and coherent (for $\alpha \geq 0.5$).[1] Its practical implementation, however, is excluded because it lacks any economic meaning. Despite ES having established a foothold in Basel III and being used as the proposed, new measure in determining the capital charge for trading book positions, the Basel Committee still supports VaR for backtesting purposes.

In light of the recent developments in backtesting methodology, there is an urgent need to revise the current system. The present work attempts to highlight the ideas and results that serve as a basis for this far-reaching conclusion.

2. Value-at-Risk versus Expected Shortfall

According to *McNeil et al.* (*2015*), the most modern approach in quantitative risk management is to identify the risk of a portfolio as a statistical quantity describing the return (loss) distribution of the portfolio over some predetermined time horizon. Examples include the variance proposed by *Markowitz* (*1952*), VaR as well as ES, the latter ones are the two measures adopted by banking regulation.

VaR is a quantile-based risk measure (*Dowd–Blake 2006*), i.e. the negative α-quantile of the return (X) distribution at a given significance level α

$$VaR_{\alpha}(X) = -q_{\alpha}(X) \qquad (1)$$

It follows from eq. (1)

$$VaR_{\alpha}(X) = -F^{\leftarrow}(\alpha), \qquad 0 < \alpha < 1 \qquad (2)$$

---

[1] See e.g. *Ziegel* (*2016*) for a proof.

$F^{\leftarrow}(\alpha)$ above is the generalised inverse of the distribution function $F(x)$ for the random variable X (see *Embrechts et al. 1999*), i.e.

$$F^{\leftarrow}(\alpha)=inf\left\{x \in \mathbb{R} \middle| F(x) \geq \alpha \right\},\qquad(3)$$

where *inf{.}* refers to the infimum, namely the greatest lower bound of the set indicated in brackets.[2]

The VaR methodology has been adopted in the form of the so-called standard model for regulation in the banking sector since 1993 with the purpose of assessing regulatory capital in order to cover market risk-related losses. The dynamic progress of the financial sector coupled with high-speed product innovations raises more and more problems and challenges for regulators to cope with. Among other things, the subprime crisis and its escalation into a global financial crisis underlined the disadvantages of VaR as a risk metric. During that period of financial turbulence, banks' losses far exceeded the regulatory capital requirements "calibrated" using VaR.

The unreliability of VaR as a risk measure has been strongly articulated by researchers as well as risk experts since the start of 2000s. A number of studies were published highlighting the "weaknesses" of VaR. Furthermore, in 2002 the Journal of Banking and Finance devoted a special issue to the problems in risk measurement theory and their applications in risk management. *Szegö* (2002) even used the provocative title "No more VaR (this is not a typo)" for an editorial note. However, the "distress signals" by representatives of risk measurement theory were not taken seriously by regulators in the banking sector until the outbreak of the subprime crisis in 2007.

The risk measure also known as Conditional Value-at-Risk (CVaR) and Tail-VaR was eventually adopted as Expected Shortfall (ES) in the Basel regulation.[3] At a given significance level α (over some predetermined time horizon) ES is a conditional expectation, namely the expected value of loss (negative return) exceeding VaR. According to *Acerbi – Tasche* (*2002*), it can be formulated as follows.

$$ES_{\alpha}(X)=VaR_{\alpha}(X)+\frac{1}{\alpha}E(X+VaR_{\alpha})_{-}{}^{4}\qquad(4)$$

[2] If the α-quantile is single-valued then VaR_{α} can be determined as the negative inverse of the distribution function at α, namely $VaR_{\alpha}(X) = -F^{-1}(\alpha)$.
[3] For details see *BCBS* (*2016*) and *Bugár – Ratting* (*2016*)
[4] The meaning of *(a)_ = −min(a,0)*. It is called the negative part of *a*.

For continuous return distributions, eq. (4) takes the form as below.

$$ES_\alpha(X) = \frac{1}{\alpha} \int_0^\alpha VaR_p(X)dp \qquad (5)$$

ES has very attractive theoretical and empirical properties compared to VaR. It accounts for losses beyond VaR, which is especially important in case of fat-tail distributions. It is also a coherent measure of risk complying with the coherence-axioms proposed by *Artzner et al.* (*1999*). By contrast, VaR is not coherent because it lacks subadditivity. Therefore, it can occur that the risk of a portfolio measured by VaR can be higher than the sum of the risk of its components. Furthermore, ES has two favourable technical properties. It is continuous with respect to the confidence level and convex with respect to the control variables, the latter being relevant in portfolio optimisation. In order to optimise within the mean-ES framework, as shown by *Rockafellar – Uryasev* (*2000*),[5] one must solve a simple linear programming problem. This makes ES very appealing in asset allocation.

Besides the lack of elicitability, there are other deficiencies with ES. Among these, as shown by *Csóka et al.* (*2007*), it is worth mentioning that in evaluating the risk of a portfolio from the perspective of general equilibrium theory, spectral risk measures (the generalisations of ES) do not take into account the relation of the examined portfolio to the market portfolio. As a consequence, the application of ES as a risk measure can lead to under- or overestimation of portfolio risk.

3. Model validation

In general, the term model validation refers to the process of verifying whether a model provides reliable predictions. In the risk industry and financial regulation, the method of model validation is referred to as backtesting. As expressed by *Acerbi – Székely* (*2017*), backtesting is a "reality check for its [i.e. the model's] output". In other words, it is a technique which is used to evaluate the accuracy of the forecasting method by comparing ex-ante predictions of some statistic to ex-post realisations of the random variable it refers to. The challenge is that neither the real value of the statistic nor its distribution can be revealed. The only thing which can be found out ex-post is just the realised value which can be regarded as a single random draw (an element of a sample).

---

[5] They used the name CVaR for ES.

In particular, in risk measurement theory the statistic considered is a point estimator of a risk measure, for example that of VaR or ES given by eq. (2) and (4), respectively. Return (*X*) is the corresponding random variable and *F*(*x*) denotes the unknown distribution function of *X*. By relying on a particular model, namely on a certain method or algorithm, with which the risk measure applied is estimated, our task is to evaluate the predictive performance based on a time series (sample) of realised returns (x_t) and compare them to the risk measure predictions (y_t). The latter means the estimated value of the relevant risk measure given at time *t* − 1 for time point *t*. This process generates another challenge, especially in the case of ES backtests. As the real value of ES cannot be observed, risk measure predictions must be tested against return outcomes. This can be referred to as "apples-with-oranges" comparison (see *Acerbi – Székely 2019*).

There are two important types of model validation. In the case of absolute validation, the predictive performance of a given predetermined model is evaluated. This is the traditional approach aimed at backtesting a single internal model by financial regulators. For that purpose, the corresponding methodology is hypothesis testing.

The other type of model validation is relative validation or model selection as termed by *Acerbi – Székely* (*2014*). This is relevant when the aim is to rank different models based on their forecasting ability. In such a situation, the property called elicitability can prove to be very beneficial.

4. Elicitability and backtestability

4.1. The relevance of elicitability in backtesting

Elicitable statistics possess the property that they minimise the expected value of a scoring function. The role of the scoring function is to quantify the discrepancy between the forecasted and realised value from the distribution. The scoring function must be strictly consistent, which means that the scoring function has a finite mean and its expected value has a unique minimum. A formal definition of elicitability can be given as follows (see *McNeil et al*. *2015*).

Let *R:X→R*(*X*) be a risk measure defined on the distribution function *F*(*x*) of the random return *X*. The risk measure *R* can also be interpreted as a statistic *T* (or a statistical functional) defined on *F*, so that *T:F→T*(*F*). The relationship above can be expressed as *R*(*X*) = *T*(*F*).

A statistic *T* of a risk measure is elicitable if there exists a strictly consistent scoring function $S: \mathbb{R} \times \mathbb{R} \rightarrow [0, \infty)$ such that for any random return *X* with finite mean

$$T(F) = argmin_{y \in \mathbb{R}} E\big(S(y,x)\big) \tag{6}$$

where *E*(.) denotes expected value. Please, note that the values of the scoring function must be non-negative.[6]

As is well-known in forecasting theory, the mean is elicitable, and the appropriate scoring function is $S(y,x) = (y - x)^2$. It can be shown that VaR is elicitable for strictly increasing distribution functions.[7] ES, however, as *Gneiting* (*2011*) has pointed out is not elicitable. It is worth mentioning that *Bellini – Bignozzi* (*2015*) have shown the same from a standpoint of financial risk measurement theory by characterising all the elicitable risk measures. Variance, which was introduced by *Markowitz* (*1952*) as a risk measure, is also not elicitable (*Lambert et al. 2008*). The above mentioned examples are summarised in *Table 1*. In the case of elicitable statistics, the relevant (canonical) scoring function is also given.

Table 1
Examples of elicitable and non-elicitable statistics

Statistic	Elicitable?	Scoring function
Mean	Yes	$(y - x)^2$
Variance	No	Does not exist
VaR$_\alpha$	Yes	$(1 - \alpha)(y - x)_{\{x < y\}} + \alpha(y - x)_{\{x \geq y\}}$ *
ES$_\alpha$	No	Does not exist

*Note: * This can also be expressed with absolute value: $|1_{\{x \leq y\}} - \alpha \,| |y - x|$.*
Source: Author's work based on Acerbi – Székely (2017)

The lack of elicitability in the case of ES has proven to be a challenge leading to some flawed conclusions. In the meantime, it has been confirmed that this property is only relevant in relative model validation (see e.g. *Davis 2014*, *Acerbi – Székely 2014*, *Fissler et al. 2015*). Therefore, one can draw the conclusion that this property has nothing to do with absolute model validation.

The importance of elicitability in backtesting stems from the fact that it makes possible to evaluate predictions (y_t) given by different risk estimation models relying only on the realisations of the random variable, namely a sequence of realised

---

[6] For further reference on the properties of scoring functions and elicitable risk measures, see *Bugár* (*2019*).
[7] See e. g. *McNeil et al.* (*2015*) for a proof.

returns (x_t). Indeed, if a risk measure is elicitable, the competing models can be ranked by computing the mean score over a certain period of time (T).

$$\bar{s} = \frac{1}{T}\sum_{t=1}^{T} S(y_t, x_t) \tag{7}$$

The competing models can be ranked based on the concept that a smaller mean score indicates a better forecasting model.

While backtesting has long been a practice in the financial industry and regulation, according to the best of our knowledge, its formal definition was only recently given in a valuable contribution by *Acerbi – Székely* (*2017*).

4.2. Defining backtestability

A statistic T of a risk measure is backtestable if there exists a backtest function $Z : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ such that for any random return X with finite mean

$$E\big(Z(y, x)\big) = 0 \text{ if and only if } y = T(F) \tag{8}$$

A further requirement is that the backtest function Z above must be strictly increasing in the prediction y, that is

$$E\big(Z(y_1, X)\big) < E\big(Z(y_2, X)\big) \quad \text{if} \quad y_1 < y_2 \tag{9}$$

It is remarkable that becktestability is defined with the intention of being capable of differentiating over- and underestimation. Indeed, positive values in eq. (8) indicate overestimation, while negative ones underestimation, respectively. The scoring functions do not possess such a directional property. It is also obvious that worse predictions are located farther away from zero.

Acerbi – Székely (2017) pointed out that becktestability implies elicitability with *y*-convex scoring function which can be determined as follows.

$$S(y, x) = \int^{y} Z(t, x)dt \tag{10}$$

given that the integral above exists.

One real challenge, however, is that the elicitability of a risk measure proved to be a necessary condition of backtestability. This means that a non-elicitable statistic cannot be backtestable. As a special case, it follows that ES is not backtestable. At this point, the adversaries of ES might relax, but this is not the end of the story...

On the other side, elicitability with a strictly *y*-convex and smooth scoring function (i.e. continuously differentiable with respect to *y*) implies backtestability. In this case, the backtest function can be identified as

$$Z(y,x) = \partial_y S(y,x) \tag{11}$$

Based on this there is a straightforward conclusion that VaR is backtestable (given that it may impose certain restrictions on the return distribution[8]). *Table 2* shows some examples of backtestable and non-backtestable statistics, and for the former ones the relevant backtest function is also given.

Table 2

Examples of backtestable and non-backtestable statistics

Statistic	Backtestable?	Backtest function
Mean	Yes	$y - x$
Variance	No	Does not exist
VaR_α	Yes	$(1 - \alpha)_{\{x < y\}} - \alpha_{\{x > y\}} + c_{\{x = y\}}{}^*$
ES_α	No	Does not exist

Note: * The constant c above must be in the interval [−α,1 − α]. For details see Acerbi − Székely (2017), p. 10.

Source: Author's work based on Acerbi − Székely (2017)

In analogy with the methodology of how elicitability can be exploited in model selection (relative model validation), one can apply the realised backtest function for ranking competing models by calculating the mean backtest value over a period *T* based on a sequence of risk measure predictions (y_t) as well as realised returns (x_t)

$$\overline{z} = \frac{1}{T}\sum_{t=1}^{T} Z(y_t, x_t) \tag{12}$$

Unlike the test formulated by eq. (7), the test above is directional, so with a positive (negative) value it is able to detect the tendency for over-(under)estimation. It is also important to emphasise that similarly to the test given by eq. (7), it also cannot be used for absolute model validation. In this case, a backtesting procedure must be hypothesis testing, as mentioned earlier. In fact, the proper backtesting procedure for absolute model validation is proposed and described by *Acerbi − Székely* (*2017*).

---

[8] See *Acerbi − Székely* (*2017*) for details.

4.3. Sharpness of a backtest

According to the definition given by *Acerbi – Székely* (*2017*), a backtestable statistic *T(F)* admits a sharp backtest if the expected value of the backtest function depends only on the predicted (*y*) and the true value of the statistic [*T(F)*]

$$E\big[Z(y,X)\big]=\phi\big(y,T(F)\big) \tag{13}$$

where ϕ is strictly increasing in *y* and strictly decreasing in *T(F)*.

As shown in *Table 2*, the backtest function of the mean is: $Z(y,x) = y - x$. Let *m* denote the prediction value and *μ* be the real value of the mean. Then $E[Z(y,X)] = m - E(X) = m - \mu$, which proves that the backtest of the mean is sharp. VaR, by contrast, does not admit a sharp backtest.[9]

The existence of a sharp backtest is of key importance because it results in predicting the real (true) value of the given statistic. Furthermore, a sharp backtest can be regarded as a direct measure of *prediction discrepancy* as it indicates the difference between the predicted and real value.

5. Ridge backtesting as the solution to the debate

5.1. Formal definition of a ridge backtest

The concept of ridge backtesting was introduced by *Acerbi – Székely* (*2017*)*.* It builds on a concept of higher order or joint elicitability (see *Acerbi – Székely 2014*, *Fissler et al. 2015* and *Fissler – Ziegel 2016*).

The formal definition (see *Acerbi – Székely 2017*) can be given as follows. A statistic T_2 admits a ridge backtest

$$Z(y_2,y_1,x)=h(y_2)-vS(y_1,x) \tag{14}$$

if it can be expressed (up to a strictly monotonic function $g:\mathbb{R}\to\mathbb{R}$) as the minimum of the expected scoring function *S* of an elicitable auxiliary statistic T_1 such that

$$\begin{aligned} T_2(F)&=g\big(min_{y\in\mathbb{R}}E\big(S(y,X)\big)\big)\\ T_1(F)&=argmin_{y\in\mathbb{R}}E\big(S(y,X)\big) \end{aligned} \tag{15}$$

In (14), $v\in\{\pm 1\}$ is the sign of g' (the derivative of *g*) and $h(x)\equiv vg^{-1}(y)$.

---

[9] See *Acerbi – Székely* (*2019*) for details.

As can be seen from (14) in the case of a ridge backtest the prediction of an auxiliary elicitable variable is involved in the backtesting procedure of the intended (in itself non-backtestable) statistic. The latter coincides the attained minimum of the expected value of the former one's scoring function.

5.2. Ridge backtest for the variance

As mentioned earlier, the variance does not comply with the requirements of a backtestable statistic, because it is not elicitable. However, there is a possibility to develop a ridge backtest utilising that the variance can be expressed as the minimum of the expected value of the mean's scoring function. Indeed, in line with (15)

$$
\begin{aligned}
V &= min_{m\in\mathbb{R}} E\left[(X-m)^2\right] \\
\mu &= argmin_{m\in\mathbb{R}} E\left[(X-m)^2\right]
\end{aligned}
$$

(16)

It follows that the backtest function of the variance can be written as

$$Z(v,m,x) = v - (x-m)^2$$

(17)

V and *μ* above denote the real values of the variance and the mean, while *v* and *m* stand for the predicted values, respectively.

The backtest of the variance is sharp up to a quadratic bias $B(m) = (m - \mu)^2$ which indicates its sensitivity to the accuracy in the prediction of the mean (see *Acerbi – Székely 2019*).

5.3. Ridge backtest for ES

Based on the representation given by *Rockafellar – Uryasev* (*2002*)

$$
\begin{aligned}
ES_\alpha &= min_{v\in\mathbb{R}} E\left[v + \frac{1}{\alpha}(X+v)_-\right] \\
VaR_\alpha &= argmin_{v\in\mathbb{R}} E\left[v + \frac{1}{\alpha}(X+v)_-\right]
\end{aligned}
$$

(18)

Acerbi – Székely (*2019*) have shown that ES admits a ridge backtest with the corresponding backtest function as follows:

$$Z(e,v,x) = e - v - \frac{1}{\alpha}(x+v)_-$$

(19)

where *α* is the given significance level, *e* and *v* denote the predicted values of ES and VaR, respectively. The test above can be applied both for model selection and absolute model validation.

The expected value of the backtest function can be expressed as

$$E\left[Z(e,v,x)\right]=e-ES_{\alpha}-B(v) \qquad (20)$$

It can be seen from eq. (20) that the ES backtest above is sharp up to the bias

$$B(v)=E\left[v+\frac{1}{\alpha}(X+v)_{-}\right]-ES_{\alpha}\geq 0 \qquad (21)$$

which is positive for any *v*, and zero if the prediction of VaR is exact, i.e. *v = VaR$_\alpha$*. *Acerbi – Székely* (*2019*) demonstrate that for continuous return distributions the bias can be approximated as the quadratic function of the discrepancy in VaR prediction[10]

$$B(v)\approx\frac{f(-VaR_{\alpha})}{2\alpha}(v-VaR_{\alpha})^{2} \qquad (22)$$

where *f* denotes the density function of the return distribution. The bias is tiny if the difference is small between the predicted and real value of VaR. It is also remarkable that *B(v)* has a prudential effect[11] in a sense that non-accurate VaR prediction (a bad VaR estimate) increases the discrepancy between the predicted and real value of ES (see eq. (20)).

In order to illustrate the relative size of the bias given in (22), we considered four distributions: the standard normal and Student's t for three different degrees of freedom.[12] Based on the calculations shown in *Table 3*, the relative size of the bias, i.e. *B(v)/|ES|* as the function of the (relative) error in the VaR prediction (*v – VaR)/v*) is shown in *Figure 1*. In calculating VaR and ES, as a significance level, *α* = 2.5 per cent has been applied. It is worth mentioning that a negative error in VaR prediction indicates underestimation, while a positive one detects overestimation.

It can be seen in the figure that the graphs are asymmetric, indicating that the relative size of the bias is more sensitive to VaR overestimation than it is to underestimation. Within a wide range of error in the VaR prediction, however, the relative size of the bias can be kept small. Indeed, between –30 per cent and 20 per cent VaR-prediction error, in all cases (except the normal distribution[13]) the relative size of the bias proves to be under 10 per cent.

---

[10] See *Acerbi – Székely* (*2017, 2019*)
[11] It is due to the non-negativity of the bias.
[12] Please note that the former can be generated from the latter for the degree of freedom converging to infinity.
[13] Even in this case it is only slightly higher than 10 per cent for VaR underestimation and around 12 per cent for its overestimation.

Table 3

VaR, ES and the multiplier of the quadratic discrepancy for the standard normal and different Student's t distributions

(α = 2.5 per cent)

Distribution	VaR$_\alpha$	ES$_\alpha$ *	Multiplier **
Student's t *(df = 2)*	−4.30	−8.40	0.22
Student's t *(df = 5)*	−2.57	−3.52	0.61
Student's t *(df = 20)*	−2.09	−2.97	1.00
Standard normal	−1.96	−2.34	1.17

*Note: * ES for t-distributions has been calculated based on the formula given by Zhang (2016). ** The multiplier is the term in (22) that is equal to f(−VaR)/2α.*

Figure 1

Relative size of the bias as a function of VaR prediction error for four different distributions

(α = 2.5 per cent)

Source: Author's own calculations based on Acerbi – Székely (2019)
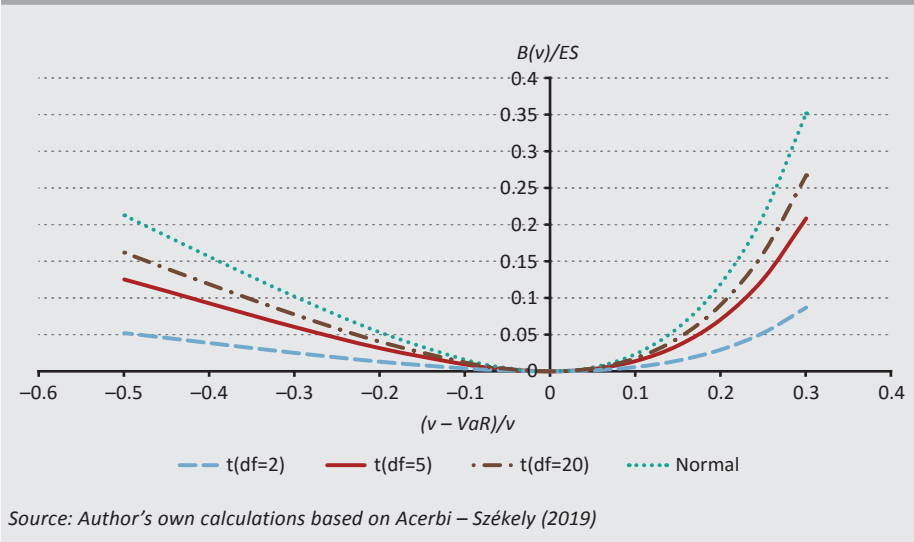
It might be surprising at first glance that the bias proves to be stronger for distributions with thinner tails. This can be regarded as an additional "clever feature" of the proposed backtest and can be explained as a penalty for applying a model with a wrongly specified tail shape (see *Acerbi – Székely* (*2019*)).

The novelty of the most recent contribution of *Acerbi – Székely* (*2019*) is the introduction of the term "realised ES". It can be formulated as

$$\widehat{ES}_\alpha = \frac{1}{T}\sum_{t=1}^{T} v_t + \frac{1}{\alpha}\left(x_t + v_t\right)_-$$

(23)

The possibility of its meaningful definition is due to the fact that the ES backtest is sharp. Indeed, the average z-score (the realised value of $E[Z(e,v,X)]$ given by eq. (20)) can be written as the difference between the average predicted and realised value of ES

$$\overline{z}_{ES_\alpha}\left(e_t,v_t,x_t\right)=\frac{1}{T}\sum_{t=1}^{T}e_t-\frac{1}{T}\sum_{t=1}^{T}v_t-\frac{1}{\alpha}\left(x_t+v_t\right)_- \qquad (24)$$

The significance of this finding is that it makes it possible to compare the predicted (ex-ante) and realised (ex-post) value of ES, and eventually – in the wording of *Acerbi – Székely* (*2019*) – to perform "apples-with-apples comparisons".

6. Summary and conclusions

This note attempts to draw attention to the existence of an effective backtesting methodology for ES. This risk measure has recently been adopted in the Basel regulation for determining the capital charge for trading book positions in internal models. While it can be regarded as a progress from the side of academics and financial experts, the adoption of ES seems to be contradictory in the sense that VaR has been kept for model-backtesting.

The non-elicitability of ES proved to be a big challenge. Furthermore, the formal definition of becktestability and revealing its connection to elicitability involves another challenge. In a recent breakthrough paper, it was pointed out by *Acerbi – Székely* (*2017*) that ES, the risk measure which already has a foothold in banking regulation, is not backtestable. In the same paper, however, the authors have shown the possibility for ES to admit a so-called ridge backtest which requires the prediction of VaR, which is elicitable, to be involved. It is built on the idea that ES is jointly elicitable with VaR shown by Fissler – *Ziegel* (2016). One appealing property of a ridge backtest is that it has limited sensitivity to the prediction of VaR and the direction of the bias is known. Furthermore, building on the concept of a sharp backtest, in the most recent contribution *Acerbi – Székely* (*2019*) have shown the possibility of introducing the notion of "realised ES" which can serve as an ex-post measure of tail risk.

These significant results pave the way for effective backtests of ES. Therefore, the methodology is ready to be applied and awaits the (re)actions of financial regulators. Promising progress in this respect is that the backtest of ES developed by Acerbi and Székely is currently under scrutiny by the risk team of the Bank of Italy, as it is being implemented to the central bank's own portfolio. The feedback seems to be very positive so far (*Cesa 2019*). The initial recognition of the method

and growing interest in the results are also indicated by a recent podcast where Acerbi was interviewed[14] on their proposed new methodology developed together with Székely.

References

Acerbi, C. – Székely, B. (2014): *Backtesting Expected Shortfall*. Risk Magazine, 27: 76–81.

Acerbi, C. – Székely, B. (2017): *General Properties of backtestable statistics. Working paper.* https://doi.org/10.2139/ssrn.2905109

Acerbi, C. – Székely, B. (2019): *The minimally biased backtest for ES*. Risk.net, 29 August. https://www.risk.net/cutting-edge/banking/6947266/the-minimally-biased-backtest-for-es. Downloaded: 6 September 2019.

Acerbi, C. – Tasche, D. (2002): *On the coherence of expected shortfall*. Journal of Banking and Finance, 26(7): 1487–1503. https://doi.org/10.1016/S0378-4266(02)00283-2

Artzner, P. – Delbaen, F. – Eber, J.M. – Heath, D. (1999): *Coherent Measures of Risk*. Mathematical Finance, 9(3): 203–228. https://doi.org/10.1111/1467-9965.00068

BCBS (2016): Basel Committee in Banking Supervision: *Minimum Capital Requirements for Market Risk.* Bank for International Settlements. 1–88.

Bellini, F. – Bignozzi, V. (2015): *Elicitable Risk Measures*. Quantitative Finance, 15(5): 725–733. https://doi.org/10.1080/14697688.2014.946955

Bugár, Gy. (2019): *Kockázati mértékek becslésének igazolása* (*Verifying risk measure estimates*). Statisztikai Szemle (Statistical Review), 97(8): 731–748. https://doi.org/10.20311/stat2019.8.hu0731

Bugár, Gy. – Ratting, A. (2016): *Revision of the quantification of market risk in the Basel III regulatory framework*. Financial and Economic Review, 15(1): 33–50. https://en-hitelintezetiszemle.mnb.hu/letoltes/2-bugar-ratting-1.pdf

Carver, L. (2013): *Mooted VaR substitute cannot be back-tested, says top quant*. Risk, 8 March.

Cesa, M. – Osborn, T. (2019): *Podcast: Acerbi on backtesting ES and FRTB's patchwork rules*. Risk.net (Cutting Edge: Views), 30 August. https://www.risk.net/risk-management/6950496/podcast-acerbi-on-backtesting-es-and-frtbs-patchwork-rules. Downloaded: 6 September 2019.

---

[14] The interview was conducted by Cesa and Osborn and is available at https://www.risk.net/risk-management/6950496/podcast-acerbi-on-backtesting-es-and-frtbs-patchwork-rules (see under the reference of *Cesa – Osborn* (*2019*)).

Cesa, M. (2019): *Backtesting expected shortfall: mission accomplished?* Risk.net (Cutting Edge: Our take), 3 September. https://www.risk.net/our-take/6963251/backtesting-expected-shortfall-mission-accomplished. Downloaded: 6 September 2019.

Csóka, P. – Herings, P.J.J. – Kóczy, L. Á. (2007): *Coherent measures of risk from a general equilibrium perspective*. Journal of Banking and Finance, 31(8): 2517–2534. https://doi.org/10.1016/j.jbankfin.2006.10.026

Davis, M. (2014): *Consistency of internal risk measure estimates*. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.2342279

Dowd, K. – Blake, D. (2006): *After VaR – The Theory, Estimation and Insurance Applications of Quantile-Based Risk Measures*. Journal of Risk and Insurance, 73(2): 193–229. https://doi.org/10.1111/j.1539-6975.2006.00171.x

Embrechts, P. – Resnick, S.I. – Samorodnitsky, G. (1999): *Extreme Value Theory as a Risk Management Tool*. North American Actuarial Journal, 3(2): 30–41. https://doi.org/10.1080/10920277.1999.10595797

Fissler, T. – Ziegel, J.F. (2016): *Higher Order Elicitability and Osband's principle*. Annals of Statistics, 444(4): 1680–1707. https://doi.org/10.1214/16-AOS1439

Fissler, T. – Ziegel, J.F. – Gneiting, T. (2015): *Expected Shortfall is jointly elicitable with Value at Risk – Implications for Backtesting*. Working paper. https://arxiv.org/pdf/1507.00244.pdf. Downloaded: 12 November 2017.

Gneiting, T. (2011): *Making and Evaluating Point Forecasts*. Journal of the American Statistical Association. 106(494): 746–762. https://doi.org/10.1198/jasa.2011.r10138

Lambert, N. – Pennock, D.M. – Shoham, Y. (2008): *Eliciting Properties of Probability Distributions*. Proceedings of the 9th ACM Conference on Electronic Commerce, EC 08. https://doi.org/10.1145/1386790.1386813

Markowitz, H.M. (1952): *Portfolio Selection*. Journal of Finance, 7(1): 77–91. https://doi.org/10.2307/2975974

McNeil, A.J. – Frey, R. – Embrechts, P. (2015): *Quantitative Risk Management – Concepts, Techniques and Tools*. Princeton University Press. Princeton and Oxford.

Rockafellar, R.T. – Uryasev, S. (2000): *Optimization of Conditional Value-at-Risk*. Journal of Risk, 2(3): 21–41. https://doi.org/10.21314/JOR.2000.038

Rockafellar, R.T. – Uryasev, S. (2002): *Conditional Value-at-Risk for General Loss Distributions*. Journal of Banking and Finance, 26(7): 1443–1471. https://doi.org/10.1016/S0378-4266(02)00271-6

Szegö, G. (2002): *No more VaR* (*this is not a typo*). Journal of Banking and Finance, 26(7): 1247–1251. https://doi.org/10.1016/S0378-4266(02)00280-7

Tasche, D. (2014): *Expected Shortfall is not elicitable. So what?* Presentation held at the University of Hannover, 23 January. https://www.stochastik.uni-hannover.de/fileadmin/institut/pdf/Talk_Tasche.pdf. Downloaded: 12 November 2017.

Zhang, S. (2016): *Two Equivalent Parametric Expected Shortfall Formulas for T-Distributions*. SSRN working paper. https://doi.org/10.2139/ssrn.2883935

Ziegel, J.F. (2016): *Coherence and Elicitability*. Mathematical Finance, 26(4): 901–918. https://doi.org/10.1111/mafi.12080