

## Article

# A Simple Traffic Light Approach to Backtesting Expected Shortfall

Nick Costanzino <sup>1,2,\*</sup> and Michael Curran <sup>3,\*</sup><sup>1</sup> Barclays Capital, 745 7th Ave, New York, NY 10019, USA<sup>2</sup> Department of Finance and Risk Engineering, NYU Tandon School of Engineering, New York, NY 11201, USA<sup>3</sup> Riskcare, New York, NY 10018, USA

\* Correspondence: Nick.Costanzino@barclays.com (N.C.); Michael.Curran@Riskcare.com (M.C.)

Received: 11 September 2017; Accepted: 04 January 2018; Published: 9 January 2018

**Abstract:** We propose a Traffic Light approach to backtesting Expected Shortfall which is completely consistent with, and analogous to, the Traffic Light approach to backtesting VaR (Value at Risk) initially proposed by the Basel Committee on Banking Supervision in their 1996 consultative document Basle Committee on Banking Supervision (1996). The approach relies on the generalized coverage test for Expected Shortfall developed in Costanzino and Curran (2015).

**Keywords:** Expected Shortfall; backtesting; Fundamental Review of the Trading Book

## 1. Introduction

Even before the initial Basel Committee consultative document ([Bank for International Settlements 2012](#)) there had been a push by both risk managers and academics to replace VaR (Value at Risk) with another risk measure that addresses VaR's deficiencies. In particular, coherent risk measures ([Acerbi and Tasche 2002a, 2002b](#); [Hall 2007](#)) satisfy the basic desired properties required by a risk measure as outlined in [Artzner et al. \(1999\)](#). Expected Shortfall (ES) is the natural choice among all coherent risk measures, and therefore there is no surprise that it has been chosen by the Basel Committee as the risk measure to replace VaR. However, unlike the case of VaR, there is no well-established backtesting framework for Expected Shortfall. Indeed, the current Basel proposal to backtest ES at the 97.5 quantile is to backtest the related VaR estimate at the 97.5 and 99 quantiles, which is a grossly insufficient test. Nevertheless, some recent backtesting methods have been proposed including, but not limited to, [Acerbi and Szekely \(2014\)](#); [Costanzino and Curran \(2015\)](#); [Du and Escanciano \(2017\)](#); [Fissler et al. \(2015\)](#); [Gordy et al. \(2017\)](#); [Kratz et al. \(2016\)](#).

The main result of this note is the development of a Traffic Light backtest for Expected Shortfall which extends the Traffic Light backtest for VaR. The test relies on the computation of critical values derived from the finite-sample distribution of the ES test statistic (9) first introduced in [Costanzino and Curran \(2015\)](#).

The note is organized as follows. In Section 2 we briefly review the VaR Traffic Light to provide context for our corresponding test for ES. In Section 3 we define the Traffic Light test for ES and compute the distribution of the finite-sample statistic from which we calculate the critical values using a numerical root-finding algorithm. Finally, in Section 4 we discuss the test and some implications.

## 2. Review of the VaR Traffic Light Test

Let  $\{t_i\}_{i=0}^N$  be a sequence of historical trading days and  $\{L_i\}_{i=1}^N$  the corresponding realized trading losses. The most basic approach to assessing the accuracy of a VaR forecast calculation for those trading days is to backtest using the VaR Coverage Test which essentially counts the number of VaR breaches. This leads to the Traffic Light approach to backtesting VaR originally proposed by the Basel Committee for Banking Supervision in [Basle Committee on Banking Supervision \(1996\)](#), which we describe below.

For each  $i = 1, \dots, N$ , let  $\text{VaR}_i(\alpha)$  denote the forecast VaR at level  $\alpha$  defined by

$$\text{VaR}(\alpha) := \inf\{z \in \mathbb{R} : F_L(z) \geq \alpha\} \quad (1)$$

where  $F_L$  is the cumulative distribution of the random loss variable  $L$ . For each trading day  $i$  we define the VaR breach indicator  $X_{\text{VaR}}^{(i)} : [0, 1] \rightarrow \{0, 1\}$  as

$$X_{\text{VaR}}^{(i)}(\alpha) := \mathbb{1}_{\{L_i \leq \text{VaR}_i(\alpha)\}} = \begin{cases} 0 & \text{if } L_i > \text{VaR}_i(\alpha) \\ 1 & \text{if } L_i \leq \text{VaR}_i(\alpha). \end{cases} \quad (2)$$

That is,  $X_{\text{VaR}}^{(i)}$  keeps track of whether a breach occurred for trading day  $i$ . Then, the total number of breaches over all  $N$  trading days, denoted by  $X_{\text{VaR}}^N : [0, 1] \rightarrow \{0, 1, 2, \dots, N\}$ , is

$$X_{\text{VaR}}^N(\alpha) := \sum_{i=1}^N \mathbb{1}_{\{L_i \leq \text{VaR}_i(\alpha)\}} \quad (3)$$

Under the null hypothesis that the VaR model is correct,  $\mathbb{E}[X_{\text{VaR}}^N(\alpha)] = N\alpha$ . Thus, for the Basel parameters  $\alpha = 1\%$  and  $N = 250$ , we expect 2.5 breaches. Of course, in any backtest it is very rare that one observes exactly 2.5 breaches (in fact impossible since  $X_{\text{VaR}}^N$  must be an integer), and thus we appeal to statistical analysis to understand the probability of obtaining significantly fewer or more breaches than would be expected if we had a correct model. For fixed  $N$  and level  $\alpha$  we define the cumulative probability  $\Psi_{\text{VaR}}^{\alpha, N}$  of obtaining  $x$  or fewer breaches as

$$\Psi_{\text{VaR}}^{\alpha, N}(x) := \mathbb{P}[X_{\text{VaR}}^N(\alpha) \leq x] \quad (4)$$

The Basel Committee on Banking Supervision proposed a Traffic Light approach to statistical significance of VaR breaches in their 1996 document [Basel Committee on Banking Supervision \(1996\)](#). Therein the Basel Committee defines three color zones through cumulative probabilities of the number of realized VaR breaches. The Green Zone is defined as the number of breaches under the null hypothesis whereby the cumulative probability of obtaining that many breaches or fewer is less than 95%. The Yellow Zone is defined as the number of breaches whereby the cumulative probability of obtaining that many breaches or fewer is greater than 95% but less than 99%. Finally, the Red Zone is defined by a cumulative probability of 99.99% or more. Thus, the boundary between the Green and Yellow zones is defined as the largest integer  $x$  such that  $\Psi_{\text{VaR}}^{\alpha, N}(x) < 0.95$  and the boundary between the Yellow and Red zones is similarly defined as the largest integer  $x$  such that  $\Psi_{\text{VaR}}^{\alpha, N}(x) < 0.9999$ .

Table 1 gives the resulting color zones for different breaches values under the VaR Basel parameters  $\alpha = 1\%$  and  $N = 250$  observations. The true Binomial Null Distribution is used to compute the Cumulative Probabilities rather than the asymptotic Normal distribution.

**Table 1.** Traffic Light zone boundaries are computed assuming  $\alpha = 1\%$  and  $N = 250$  observations.

Basel Traffic Light Approach to VaR		
Zone	Breach Value	Cumulative Probability
Green	0	8.11%
	1	28.58%
	2	54.32%
	3	75.81%
	4	89.22%

Table 1. Cont.

Basel Traffic Light Approach to VaR		
Zone	Breach Value	Cumulative Probability
Yellow	5	95.88%
	6	98.63%
	7	99.60%
	8	99.89%
	9	99.97%
Red	more than 10	99.99%

### 3. Derivation of the Expected Shortfall Traffic Light Test

We now define a Traffic Light approach to backtesting Expected Shortfall based on the Coverage Test in Costanzino and Curran (2015). The test relies on an appropriate extension of the VaR breach indicator (2) to the case of ES. The resulting new breach indicator (6) takes into account the severity of the breach (i.e., losses beyond the VaR level) and is a continuous variable rather than discrete.

We begin the derivation by defining Expected Shortfall as

$$ES(\alpha) := \frac{1}{\alpha} \int_0^\alpha VaR(p) dp \quad (5)$$

In analogy to  $X_{VaR}^{(i)}$  (2), we define the ES generalized breach indicator  $X_{ES}^{(i)} : [0, 1] \rightarrow [0, 1]$  by

$$X_{ES}^{(i)}(\alpha) := \frac{1}{\alpha} \int_0^\alpha \mathbb{1}_{\{L_i \leq VaR_i(p)\}} dp \quad (6)$$

$$= \underbrace{\left(1 - \frac{F_L(L_i)}{\alpha}\right)}_{\text{severity of the breach}} \mathbb{1}_{\{L_i \leq VaR_i(\alpha)\}} \quad (7)$$

$$= \theta^{(i)}(\alpha) \cdot X_{VaR}^{(i)} \quad (8)$$

where we have used (2) and have set  $\theta^{(i)}(\alpha) = 1 - F_L(L_i)/\alpha$ , where  $F_L$  is the cumulative distribution implicitly defined in (1). We note that compared to  $X_{VaR}^{(i)}$  (2),  $X_{ES}^{(i)}$  (6) has an extra term  $F(L_i)/\alpha$  which determines the severity of the breach. That is, suppose  $L_i = VaR_i$ . Then  $F(L_i)/\alpha = 1$  so  $X_{ES}^{(i)} = 0$  whereas  $X_{VaR}^{(i)} = 1$ . On the other hand suppose  $L_i$  is very negative. Then,  $F(L_i) = 0$  so that  $X_{ES}^{(i)} = 1$  and similarly  $X_{VaR}^{(i)} = 1$ . Thus,  $X_{ES}^{(i)}$  keeps track of whether a breach happened on trading day  $i$  as well as the severity. Then, the total severity of breaches over all  $N$  trading days, denoted by  $X_{ES}^N : [0, 1] \rightarrow [0, N]$ , is

$$\begin{aligned} X_{ES}^N(\alpha) &:= \sum_{i=1}^N \frac{1}{\alpha} \int_0^\alpha \mathbb{1}_{\{L_i \leq VaR_i(p)\}} dp \\ &= \sum_{i=1}^N \left(1 - \frac{F(L_i)}{\alpha}\right) \mathbb{1}_{\{L_i \leq VaR_i(\alpha)\}} \\ &= \sum_{i=1}^N \theta^{(i)}(\alpha) X_{VaR}^{(i)}(\alpha) \end{aligned} \quad (9)$$

For fixed  $N$  and level  $\alpha$  we define the cumulative probability  $\Psi_{ES}^{\alpha, N}$  of obtaining  $x$  or fewer breaches as

$$\Psi_{ES}^{\alpha, N}(x) := \mathbb{P}[X_{ES}^N(\alpha) \leq x] \quad (10)$$

Therefore, for any quantile  $q$ , we can compute the corresponding Generalized Breach Value  $x$  by inverting the equation

$$\sup_{x \in [0, \infty)} \Psi_{\text{ES}}^{\alpha, N}(x) < q \quad (11)$$

Note that in the case of the VaR Traffic Light Test (see Table 1), it makes sense to compute the quantiles for different breach values (i.e., 1, 2, ..., 10). For Expected Shortfall, the breach indicator is a continuous variable and it no longer makes sense to choose the breach value and compute an associated quantile. Rather, we choose the quantile and then invert to obtain the corresponding breach value. In particular, we borrow the color zone boundaries from the VaR Traffic Light Test, which yield a Green Zone if  $q < 0.95$ , Yellow Zone if  $0.95 \leq q < 0.9999$ , and Red Zone if  $0.9999 \leq q$ ; i.e.,

$$\text{Boundary}_{\text{GY}} := \sup_{x \in [0, \infty)} \Psi_{\text{ES}}^{\alpha, N}(x) < 0.95 \quad (12)$$

and the boundary between the Yellow and Red zones is given by

$$\text{Boundary}_{\text{YR}} := \sup_{x \in [0, \infty)} \Psi_{\text{ES}}^{\alpha, N}(x) < 0.9999 \quad (13)$$

To compute these boundaries, and other values of  $x$  one needs to compute the distribution of the test statistic  $X_{\text{ES}}^N(\alpha)$  under the null-hypothesis  $\mathfrak{H}_0$  given by

$$\mathfrak{H}_0 : \{X_{\text{ES}}^{(i)}\}_{i=1}^N \text{ i.i.d. } \forall i \neq j, \text{ and } \mathbb{P}[L_i \leq \text{VaR}_i(p)] = p \quad \forall p \in [0, \alpha]. \quad (14)$$

A similar argument as in Costanzino and Curran (2015) shows that for any  $\alpha \in (0, 1)$ ,

$$\sqrt{N} \left( X_{\text{ES}}^N(\alpha) - \mu_{\text{ES}} \right) \xrightarrow[N \rightarrow \infty]{} \mathcal{D} \longrightarrow \mathcal{N}(0, \sigma_{\text{ES}}^2) \quad (15)$$

where

$$\mu_{\text{ES}} = \frac{1}{2} \alpha N \quad (16)$$

$$\sigma_{\text{ES}}^2 = \alpha \left( \frac{4 - 3\alpha}{12} \right) \quad (17)$$

and hence

$$\lim_{N \rightarrow \infty} X_{\text{ES}}^N(\alpha) \sim \mathcal{N}(\mu_{\text{ES}}, \sigma_{\text{ES}}^2) \quad (18)$$

Hence, as a crude approximation, we can compute (12) and (13) using the asymptotic test distribution (18) to obtain  $\text{Boundary}_{\text{GY}}^{\text{approx}} = 5.4768$  and  $\text{Boundary}_{\text{YR}}^{\text{approx}} = 9.2229$ . These values are approximate since they use the asymptotic distribution  $\Psi_{\text{ES}}^{\alpha, \infty}(x)$  of the test statistic rather than the finite-sample one  $\Psi_{\text{ES}}^{\alpha, N}(x)$ . We now derive the finite-sample distribution  $\Psi_{\text{ES}}^{\alpha, N}(x)$  and use a numerical root-finding procedure to accurately estimate the critical values.

The derivation of the ES Traffic Light test relies on the computation of the finite-sample cumulative distribution of the test statistic  $X_{\text{ES}}^N(\alpha)$  (9). A key observation in the derivation is that under the null-hypothesis, the distribution of  $X_{\text{ES}}^{(i)}(\alpha)$  conditional on a breach having occurred is uniform in the  $\alpha$ -tail, and thus using the law of total probability we have

$$\begin{aligned}\mathbb{P}[X_{ES}^N(\alpha) \leq x] &= \sum_{n=1}^{\infty} \mathbb{P}\left[X_{ES}^N(\alpha) \leq x \mid X_{VaR}^N(\alpha) = n\right] \cdot \mathbb{P}[X_{VaR}^N(\alpha) = n] \\ &= \sum_{n=1}^{\infty} \mathcal{I}_n(x) \cdot \mathfrak{B}(n, N, \alpha)\end{aligned}\quad (19)$$

where  $\mathcal{I}_n(x)$  is the Irwin–Hall distribution (c.f. [Hall 1927](#); [Irwin 1927](#); [Marengo et al. 2017](#)) defined by

$$\mathcal{I}_n(x) = \frac{1}{2(n-1)!} \sum_{k=0}^n (-1)^k \binom{n}{k} (x-k)^{n-1} \text{sgn}(x-k) \quad (20)$$

and  $\mathfrak{B}(n, N, \alpha)$  binomial probability mass function

$$\mathfrak{B}(n, N, \alpha) = \binom{n}{N} \alpha^n (1-\alpha)^{N-n} \quad (21)$$

We then use this probability calculation (19) and a root-finding algorithm to solve the equation

$$\mathbb{P}[X_{ES}^N(\alpha) \leq \text{Boundary}] = q \quad (22)$$

for Boundary where  $q$  is the appropriate quantile level. In particular assuming the Basel parameters  $\alpha = 2.5\%$  and  $N = 250$ , then for  $q = q_{GY} = 0.05$  and  $q = q_{YR} = 0.0001$  we obtain

$$\text{Boundary}_{GY} = 5.7049 \quad (23)$$

$$\text{Boundary}_{YR} = 9.8833 \quad (24)$$

Table 2 gives the resulting quantiles and color zones for different breach values under the ES Basel parameters  $\alpha = 2.5\%$  and  $N = 250$  observations where the cumulative probabilities were computed using (19). Of particular note is the breach values and cumulative probabilities for Expected Shortfall at the 97.5 quantile (i.e.,  $\alpha = 2.5\%$ ) are very similar to the VaR values at the 99 quantile (i.e.,  $\alpha = 1\%$ ). In addition, the finite-sample Breach Value at the 50th quantile (3.0276) is very similar to the asymptotic Breach Value at the 50th quantile ( $\frac{\alpha}{2}N = 250 \times 0.025/2 = 3.125$ ). Furthermore, note that in the case of Expected Shortfall, the breach values are continuous, and therefore infinitesimally small changes in breach value may result in a change in the color zone.

**Table 2.** Expected Shortfall Traffic Light zone boundaries are computed assuming  $\alpha = 2.5\%$  and  $N = 250$  observations.

Traffic Light Approach to Expected Shortfall		
Zone	Generalized Breach Value	Cumulative Probability
Green	0	0.18%
	1.3929	10%
	2.1131	25%
	3.0276	50%
	4.0520	75%
	5.0622	90%
	5.7049	95%
Yellow	5.7049	95%
	6.9844	99%
	8.5285	99.9%
	9.8833	99.99%
Red	more than 9.8833	99.99%

#### 4. Discussion

First, the values and quantiles for VaR at  $\alpha = 1\%$  are similar to the values and quantiles for ES at  $\alpha = 2.5\%$ . This happens because there are more VaR breaches at  $\alpha = 2.5\%$  than at  $\alpha = 1\%$ , but the severity of the breach in ES is smaller than unity so these two mechanisms average each other out.

We also note that along with color zones, the Basel document (Basle Committee on Banking Supervision 1996) defines market risk capital multipliers based on the cumulative probability  $C_{\text{VaR}}$  of the number of realized exceptions,  $X_{\text{VaR}}$ . In particular, a multiplier  $k_{\text{VaR}}$  ranging from 0 to 1 is given depending on the number of breaches; i.e.,  $k_{\text{VaR}} = f_{\text{VaR}}(C_{\text{VaR}})$  for some function  $f_{\text{VaR}}$ . The same can obviously be done for Expected Shortfall; i.e.,  $k_{\text{ES}} = f_{\text{ES}}(C_{\text{ES}})$  for some function  $f_{\text{ES}}$ . However, the continuous nature of the breach values from (9) leads to the need for  $k_{\text{ES}}$  to be a continuous function so as to avoid the case where small changes in breach value give rise to large changes in multiplier.

#### 5. Conclusions

By defining an appropriate breach statistic (6) that measures the severity of each breach and using the results in (Costanzino and Curran 2015; Clift et al. 2015), we propose a Traffic Light test for Expected Shortfall using the finite-sample distribution of the test statistic under the null hypothesis. The test itself as well as the actual values of the zone boundaries are analogous to the Basel Traffic Light test for VaR.

**Acknowledgments:** The authors would like to thank the three anonymous referees for helpful suggestions, as well as Marco Naldi (Barclays) for a careful reading of the manuscript.

**Author Contributions:** The two authors contribute equally to this article.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

- Acerbi, Carlo, and Balazs Szekely. 2014. Back-Testing Expected Shortfall. *Risk Magazine*, November.
- Acerbi, Carlo, and Dirk Tasche. 2002a. Expected Shortfall: A natural coherent alternative to Value at Risk. *Economic Notes* 31: 379–88.
- Acerbi, Carlo, and Dirk Tasche. 2002b. On the Coherence of Expected Shortfall. *Journal of Banking & Finance* 26: 1487–503.
- Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9: 203–28.
- Basle Committee on Banking Supervision. 1996. Supervisory Framework For The Use of Back-Testing in Conjunction With The Internal Models Approach to Market Risk Capital Requirements. Available online: [www.bis.org/publ/bcbs22.htm](http://www.bis.org/publ/bcbs22.htm) (accessed on 7 January 2018).
- Costanzino, Nick, and Mike Curran. 2015. Backtesting General Spectral Risk Measures with Application to Expected Shortfall. Available online: <https://ssrn.com/abstract=2514403> (accessed on 7 January 2018).
- Clift, Simon, Nick Costanzino, and Mike Curran. 2015. Comparison of Backtesting Methods for Expected Shortfall. Preprint. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2618345](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2618345) (accessed on 7 January 2018).
- Du, Zaichao, and Juan Carlos Escanciano. 2017. Backtesting Expected Shortfall: Accounting for Tail Risk. *Management Science* 63: 940–58.
- Fissler, Tobias, Johanna F. Ziegel, and Tilmann Gneiting. 2015. Expected Shortfall is Jointly Elicitable with Value at Risk—Implications for Backtesting. *Risk Magazine*, December 21.
- Bank for International Settlements. 2012. *Fundamental Review of the Trading Book: A Revised Market Risk Framework*. Consultative Document. Basel: Basel Committee on Banking Supervision.
- Gordy, Michael B., Hsiao Yen Lok, and Alexander J. McNeil. 2017. Spectral backtests of forecast distributions with application to risk management. *arXiv Preprint arXiv:1708.01489*.
- Hall, Philip. 1927. The distribution of means for samples of size  $N$  drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika* 19: 240–45.
- Hull, John 2007. VaR versus Expected Shortfall. *Risk Magazine*, March 1.

- Irwin, Joseph Oscar. 1927. On the frequency distribution of the means of samples from a population having any law of frequency with finite moments, with special reference to Pearson's Type II. *Biometrika* 19: 225–39.
- Kratz, Marie, Yen H. Lok, and Alexander J. McNeil. 2016. Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2898688](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2898688) (accessed on 7 January 2018).
- Marengo, James E., David L. Farnsworth, and Lucas Stefanic. 2017. A Geometric Derivation of the Irwin-Hall Distribution. *International Journal of Mathematics and Mathematical Sciences* 2017: 3571419.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).