Backtesting Lambda Value at Risk

Jacopo Corbetta

CERMICS, École des Ponts , UPE, Champs sur Marne, France.

Zeliade Systems, 56 rue Jean-Jacques Rousseau, Paris, France.

and

Ilaria Peri

Department of Economics, Mathematics and Statistics, Birkbeck University of London, England

June 17, 2017

Abstract

A new risk measure, Lambda value at risk (ΛVaR), has been recently proposed as a generalization of Value at risk (VaR). ΛVaR appears attractive for its potential ability to solve several problems of VaR. This paper provides the first study on the backtesting of ΛVaR. We propose three nonparametric tests which exploit different features. Two tests are based on simple results of probability theory. One test is unilateral and is more suitable for small samples of observations. A second test is bilateral and provides an asymptotic result. A third test is based on simulations and allows for a more accurate comparison among $\Lambda VaRs$ computed with different assumptions on the asset return distribution. Finally, we perform a backtesting exercise that confirms a higher performance of ΛVaR in respect to VaR especially when it is estimated with distributions that better capture tail behaviour.

Keywords: backtesting, hypothesis testing, model validation, risk management.
JEL Codes: C12, C52, G32

1 Introduction

Risk measurement and its backtesting are matter of primary concern to financial industry. Value at risk (VaR) has become the most widely used risk measure. Despite its popularity, after the recent financial crisis, VaR has been extensively criticized by academics and risk managers. Among these critics, we recall the inability to capture the tail risk and the lack of reactivity to market fluctuations. Thus, the suggestion of the Basel Committee, in the consultative document Fundamental review of the trading book (2013), is to consider alternative risk measures that are able to overcome the VaR's weaknesses.

A new risk measure, Lambda Value at Risk (ΛVaR), has been introduced by a theoretical point of view by Frittelli, Maggis, and Peri (2014). ΛVaR is a generalization of the VaR at confidence level λ. Specifically, ΛVaR considers a function Λ instead of a constant confidence level λ, where Λ is a function of the losses. Formally, given a monotone and right continuous function $\Lambda : \mathbb{R} \to (0,1)$, the ΛVaR of the asset return X is a map that associates to its cumulative distribution function $F(x) = P(X \leq x)$ the number:

$$\Lambda VaR = -\inf \{x \in \mathbb{R} \mid F(x) > \Lambda(x)\} \ . \tag{1}$$

This new risk measure appears to be attractive for its potential ability to solve several problems of VaR. First of all, it seems to be flexible enough to discriminate the risk among return distributions with different tail behavior, by assigning more risk to heavy-tailed return distributions and less in the opposite case. In addition, ΛVaR may allow for a rapid changing of the interval of confidence when the market conditions change.

Recently, Hitaj, Mateus, and Peri (2015) proposed a methodology for computing

ΛVaR and a first attempt of backtesting based on the hypothesis testing framework by Kupiec (1995). In this study, the accuracy of the ΛVaR model is evaluated by considering the maximum of the Λ function as confidence level. However, the level of coverage provided by the ΛVaR model may not be constant at any time; hence, this method misses to assess the actual ΛVaR performance.

The objective of this paper is to propose the first theoretical framework for the backtesting of ΛVaR. We present three backtesting methodologies which exploit different features and may be used with different aims. Our tests evaluate if ΛVaR provides an accurate level of coverage, this means that the probability of a violation occurring *ex-post* actually coincides with the one predicted by the model. In respect to the hypothesis test proposed in Hitaj, Mateus, and Peri (2015), we consider a null hypothesis which better evaluates the benefits introduced by the ΛVaR flexibility. Our tests can be easily extended to VaR allowing for a proper comparison among the two risk measures.

Two of these tests are based on simple test statistics whose distribution is obtained by applying results of probability theory. The first test is unilateral and provides more precise results for shorter backtesting time windows (e.g. 250 observations). The second test is bilateral and provides an asymptotic result that makes it more suitable for larger samples of observations.

We propose a third test that is inspired to the approach used by Acerbi and Szekely (2014) for the Expected Shortfall backtesting. Here, the distribution of the test statistic is obtained by Monte Carlo simulations. This test allows to better evaluate the impact of the assumption on the model generating data and compare different choices on the asset return distribution.

Finally, we conduct an empirical analysis where we experiment and compare the results of our backtesting proposals for ΛVaR, computed using the same dynamic benchmark approach proposed by Hitaj, Mateus, and Peri (2015). The backtesting exercise has been performed along six different time windows throughout all the global financial crisis (2006-2011).

The paper is structured as follows: Section 2 presents the VaR and ΛVaR models; Section 3 introduces our backtesting proposals; Section 4 describes and shows the results of the empirical analysis; Appendix collects the proofs.

2 VaR and ΛVaR models

Let us consider a probability space $(\Omega, (\mathcal{F}_t)_T, \mathbb{P}_t)$, where the sigma algebra $\mathcal{F}_t$ represents the information at time t. We assume that X is the random variable of the returns of an asset distributed along a real (unknown) distribution F_t, i.e. $F_t(x) := \mathbb{P}_t(X_t < x)$, and it is forecasted by a model predictive distribution P_t conditional to previous information, i.e. $P_t(x) = \mathbb{P}_t(X_t \leq x|\mathcal{F}_{t-1})$.

We can measure the risk of the asset return X using the classical VaR, by attributing to X at time t the following value:

$$VaR_t = -\inf \{x \in \mathbb{R} \mid P_t(x) > \lambda\} \ . \tag{2}$$

The objective of this study is the alternative risk measure proposed by Frittelli, Maggis, and Peri (2014), ΛVaR, that attributes to X at time t the following value:

$$\Lambda VaR_t = -\inf\left\{x \in \mathbb{R} \mid P_t(x) > \Lambda_t(x)\right\} . \tag{3}$$

where Λ_t is a monotone function that maps $x \in \mathbb{R}$ in (λ_m, λ_M) with $\lambda_m > 0$ and $\lambda_M < 1$. When Λ_t is constant and equal to $\lambda \in (0,1)$ for any x, ΛVaR coincides with VaR at confidence level λ. The interesting feature of ΛVaR is the sensitivity to tail risk, in particular, it is able to discriminate the risk of assets having the same VaR at some level λ but different tail behavior. Thus, ΛVaR may allow to enhance the capital requirement in case of expected greater losses.

Hitaj, Mateus, and Peri (2015) proposed a method to compute the Λ function that is called dynamic benchmark approach. Here, the Λ function is taken as proxy of the tails of the market return distribution. This feature allows ΛVaR to assess the different asset reactions in respect to the market by detecting different confidence levels. This approach is also dynamic since Λ is re-estimated at each time t according the information in $t-1$. In this way, ΛVaR incorporates the recent market fluctuations and adjusts the confidence level according the different asset reactions.

The authors proposed different models to compute ΛVaR. One proposal is to obtain Λ by linear interpolation of n points (π_i, λ_i) for any $\pi_1 \leq x < \pi_n$ and fix Λ constantly equal to the lower (upper) bound for any $x \leq \pi_1$ and to the upper (lower) bound for any $x \geq \pi_n$ in the increasing (decreasing) case. In their empirical analysis, the authors chose 4 points ($n = 4$). In particular, on the probability axis, they set the Λ lower bound $\lambda_m = 0.001$, the upper bound $\lambda_M = 0.01$ and the others λ_i values, with $i = 2, .., 3$, by an equipartition of the interval $(0, \lambda_M]$. On the losses axis, they fix 4 points π_i equal to order statistics of the return distribution of some selected market benchmarks. Specifically, π_1 is equal to the minimum of all benchmark returns: $\pi_1 = \min x_{t,j}$ where $x_{t,j}$ is the realized return of the j-th benchmark, for $t = 1, .., T$ and T is the time horizon (i.e. number of days in the rolling window), and for $j = 1, \ldots, B$ and B is the number of benchmarks; π_2, π_3, and π_4 are equal to the maximum, mean, and minimum of the benchmarks' $\lambda\%$-VaR, respectively.

In the next section, we will recall the first attempt of backtesting for ΛVaR, explain its limit and introduce our hypothesis test proposals.

3 VaR and ΛVaR backtesting models

The Basel Committee on Banking Supervision (1996) refers to backtesting as the process of "comparing daily profits and losses with model-generated risk measures to gauge the quality and accuracy of risk measurement systems". A violation occurs when the risk measure estimate is not able to cover the realized return (profit and loss, P&L). In the same Basel 2 Accord, the Commitee has also set up the first regulatory backtesting framework for the VaR measure, known as traffic light approach. This procedure monitors the 1% VaR violations over the last 250 days. Afterwards, many alternative proposals have been introduced in the literature for VaR; we refer to Campbell (2005), Christoffersen (2010), and Berkowitz et al. (2011) for a detailed review.

Let us denote with x_t the realization of the asset return X at time t. In order to perform the backtesting of a risk measure, we need to construct the sequence of random variables representing the violations, $\{I_t\}_{t=1}^{T}$, across T days, as follows:

$$I_t = \begin{cases} 1 & \text{if } x_t < y_t \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where y_t is the return forecasted by the risk measure. The hit sequence is equal to 1 on day t if the realized returns on that day, x_t, is smaller than the value y_t predicted by the risk measure at time $t-1$ for the day t, i.e. ΛVaR_t or VaR_t. If y_t is not exceeded (or violated), then the hit sequence returns a 0. We observe that I_t is a random variable that follows a Bernoulli distribution, that is:

$$I_t \sim B(\lambda_t) \tag{5}$$

where λ_t is the probability of having an exception at time t.

In the following, we focus on testing the unconditional coverage property of the risk measures that assumes the independence of the violations I_t. A common practice in the industry is testing the independence by visual inspection of the cluster of the exceptions (see Acerbi and Szekely (2014), section 1). We conduct an empirical analysis with the available data which shows that ΛVaR clusters the exceptions considerably less than VaR and suggests a higher level of independence of the ΛVaR exceptions (as shown in Figure (1)). The reason behind might be that ΛVaR is recalculated at each time t incorporating the recent market movements and, in this way, it may avoid sequential violations. However, in order to have a complete assessment of the accuracy of a risk measure, a specific test of independence is required. In the case of ΛVaR, one cannot rely on the immediate extension of the VaR framework since the exceptions are not identically distributed. This requires a more complex analysis that we leave for a future study.

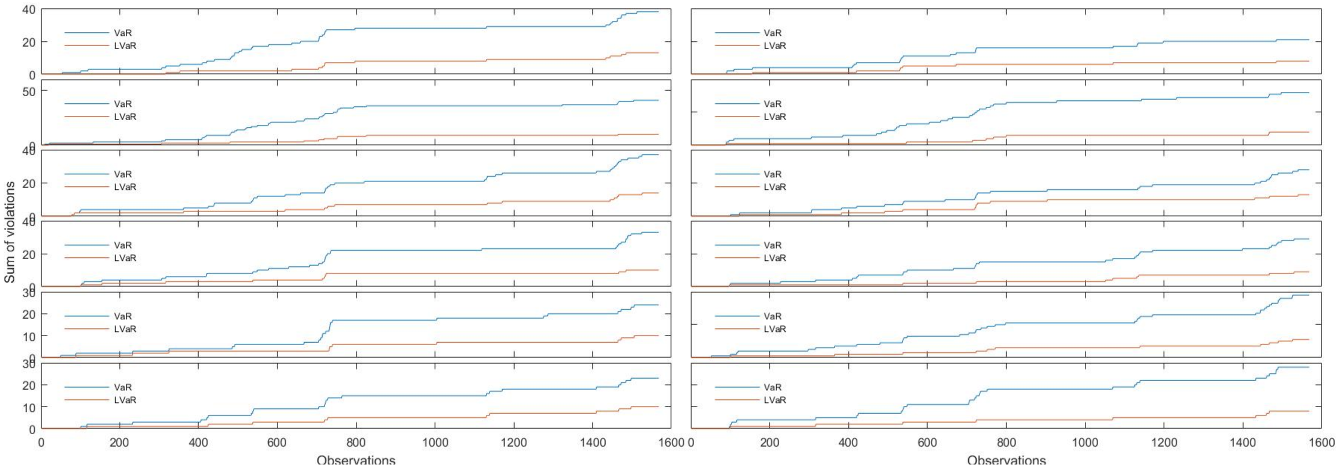

Figure 1: Time evolution of the sum of violations for $1\%VaR$ and ΛVaR. The table shows the evolution over the global financial crisis of the sum of violations of the $1\%VaR$ and the increasing ΛVaR model.

The first theoretical proposal for the backtesting of VaR is given by Kupiec (1995), where the author considers the following null and alternative hypothesis:

$$\begin{aligned} &H_0^K : \lambda_t \leq (=)\lambda^0 \quad \text{for any } t \\ &H_1^K : \lambda_t > \lambda^0 \quad \text{for some } t \text{ and equal otherwise} \end{aligned} \tag{6}$$

where λ^0 is the VaR confidence level. The VaR at level λ^0 is accepted if the frequency of the exceptions does not exceed the confidence level λ^0 for any t.

Recently, Hitaj, Mateus, and Peri (2015) have proposed a backtesting method for ΛVaR by adapting the classical Kupiec test for VaR. They consider the following null and alternative hypothesis:

$$\begin{aligned} &H_0^K : \lambda_t \leq \max(\Lambda) \quad \text{for any } t \\ &H_1^K : \lambda_t > \max(\Lambda) \quad \text{for some } t \text{ and equal otherwise} \end{aligned} \tag{7}$$

Substantially, ΛVaR is accepted if the frequency of violations is less than $\max(\Lambda)$. This is an unilateral hypothesis test that can be conducted by using the same log-likelihood ratio and critical value of the VaR test. This approach permits to verify if the coverage objective given by the Λ maximum has been reached, however, it does not allow to evaluate the accuracy of ΛVaR at any time t.

Indeed, if the ΛVaR model is correct, at time t we should be expecting that the hit sequence assumes value 1 with probability

$$\lambda_t^0 = \Lambda_t(-\Lambda VaR_t) \tag{8}$$

and 0 with probability $1 - \lambda_t^0$. This intuition is correct if both Λ_t and P_t are continuous. In case this does not occur, we have $\lambda_t^0 = P_t(-\Lambda VaR_t)$.

As a consequence, the random variables I_t of the violations for ΛVaR are not identically distributed, which implies that usual likelihood backtesting framework (POF by Kupiec 1995 , TUFF by Christoffersen 2010 etc.) cannot be directly applied.

Hence, if ΛVaR is correct, the null hypothesis should be:

$$H_0 : \lambda_t = \lambda_t^0 \text{ for any } t \tag{9}$$

while the alternative hypothesis, either:

$$H_1 : \lambda_t \neq \lambda_t^0 \text{ for some } t \tag{10}$$

in case of a bilateral test, or:

$$H_1 : \lambda_t > \lambda_t^0 \text{ for some } t \text{ and equal otherwise} \tag{11}$$

in case of an unilateral test where we reject in presence of risk under-estimation.

The null hypothesis in (9) allows to evaluate if ΛVaR guarantees the level of coverage predicted by the λ_t^0 parameter. In this way, we are able to assess the correctness of ΛVaR more precisely than Hitaj, Mateus, and Peri (2015). Notice that a rejection of H_0^K in (7) implies a rejection of H_0 in (9). Observe also that these hypothesis tests are also valid for VaR at confidence level λ^0 by fixing $\lambda_t^0 = \lambda^0$ for any t.

In order to test the accuracy of the ΛVaR model, we propose three test statistics. The distribution of the first two test statistics is obtained by exploiting simple results of probability theory. In particular, the second test provides an asymptotic result, hence it is more suitable for larger samples of observations (i.e. time horizon larger than 500).

We propose also a third test that is more useful to check if ΛVaR has been estimated with the correct distribution function, P_t. Here, the correctness of the null hypothesis is evaluated by a simulation exercise.

We suggest that the first two tests are used for an initial validation of the ΛVaR model, while the third test is used as second step for selecting the best choice of estimation for the asset return distribution.

3.1 Test 1

We set the null and the alternative hypothesis as in (9) and (11), respectively. We construct this first test by defining the test statistic Z_1 equal to the number of violations over the time horizon T, as follows:

$$Z_1 := \sum_{t=1}^{T} I_t \tag{12}$$

The distribution of Z_1 is obtained by applying classical results of probability theory. If the violations I_t independently occurs, the sum of independent Bernoulli with different mean follows a Poisson Binomial distribution (λ_t), thus we have that under H_0:

$$Z_1 \sim \text{Poiss.Bin}(\{\lambda_t^0\}). \tag{13}$$

This test is in principle a bilateral test, with critical region: $C = \left\{z_1 : z_1 < q_{Z_1}(\frac{\alpha}{2})\right\} \cup \left\{z_1 : z_1 \geq q_{Z_1}(1 - \frac{\alpha}{2})\right\}$, where α denotes the significance level of the test (i.e. 1 type error) and q_{Z_1} is the quantile of the Z_1 distribution under H_0, i.e. P_{Z_1}. However, in the backtesting practice, this test can be treated as unilateral, where the critical region is given by:

$$C_{Z_1} = \{z_1 : z_1 \geq q_{Z_1}(1 - \alpha)\} = \{z_1 : P_{Z_1}(z_1) > 1 - \alpha\} \tag{14}$$

Indeed, the probability that z_1 falls in the left side of the critical region C is null, since $q_{Z_1}(\frac{\alpha}{2})$ is zero any time the following relation is satisfied: $(1 - \max(\lambda_t))^T > \alpha/2$. This is typical for usual test significance levels ($\alpha = 10\%$ or lower), usual time horizon $T = 250$ and 1%-VaR or 1%-ΛVaR (since $\lambda_t \leq 0.01$).

This test represents an extension of the traffic light approach by Basel Committee on Banking Supervision (1996) to ΛVaR with two bands instead of three. In particular, for VaR at confidence level λ^0, under H_0 we have:

$$Z_1 \sim \text{Bin}(T, \lambda^0)$$

that is Z_1 follows a Binomial distribution. In the empirical analysis we fix $\alpha = 10\%$ and we compare the results with VaR.

3.2 Test 2

We propose a second test statistic that is founded on a result of probability theory known as Lyapunov theorem. We set the null and the alternative hypothesis as in (9) and (10), respectively. We propose another test statistic defined as follows:

$$Z_2 := \frac{\sum_{t=1}^{T}(I_t - \lambda_t^0)}{\sqrt{\sum_1^T \lambda_t^0(1-\lambda_t^0)}} \tag{15}$$

Under H_0, Z_2 is asymptotically distributed as a Standard Normal, formally:

$$Z_2 \xrightarrow{d} N(0,1) \tag{16}$$

This result follows from the application of Lemma 2 and the Lyapunov's theorem (see Appendix for details).

We remark that this is a bilateral test. Thus, we reject the null hypothesis H_0 if the realization z_2 of the test statistic stays in the following critical region:

$$C_{Z_2} := \left\{ z_2 : z_2(x) < q_{Z_2}\left(\frac{\alpha}{2}\right) \right\} \cup \left\{ z_2 : z_2(x) > q_{Z_2}\left(1 - \frac{\alpha}{2}\right) \right\} \tag{17}$$

where α is the significance level of the test, and q_{Z_2} is the quantile function of the Standard Normal distribution P_{Z_2}.

Also for this test, in the empirical analysis, we fix $\alpha = 10\%$ and we compare the results with VaR.

3.3 Test 3

The third test is inspired by Acerbi and Szekely (2014) and focused on another aspect. The aim of this test is to directly verify if ΛVaR has been estimated under the correct assumption on the distribution P_t of the returns. To this purpose we build a test statistic, Z_3, and we proceed by simulating its distribution using the same assumption as for the asset return distribution in the risk measure computation.

We set the null and the alternative hypothesis as in (9) and (11), respectively, and we define Z_3 as follows:

$$Z_3 := \frac{1}{T}\sum_{t=1}^{T}(\lambda_t^0 - I_t) = \frac{1}{T}\sum_{t=1}^{T}\lambda_t^0 - \frac{1}{T}\sum_{t=1}^{T} I_t \tag{18}$$

We observe that under H_0, we have $\mathrm{E}[Z_3] = 0$, while under H_1, $\mathrm{E}[Z_3] < 0$ for ΛVaR (see Proposition (3) in Appendix). So, if the model is correct the realized value z_3 is expected to be zero. On the other hand, a negative z_3 is a signal that the model estimation does not allow for covering the risk.

Under H_0 the distribution of Z_3 depends on the assumption for the distribution P_t of the asset returns. Hence, we perform the test by simulating M scenarios of the distribution P_t of the returns at each time t, with $t = 1, \ldots, T$. In this way, we obtain at time T the distribution P_{Z_3} of the test statistic under H_0. In order to construct the critical region we need to study the behavior of P_{Z_3} when the distribution of the returns changes from P to F. Let us compute P_{Z_3}:

$$
\begin{aligned}
P_{Z_3} = \mathbb{P}\left(Z_3 \leq z\right) &= \mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}(\lambda_t^0 - I_t) \leq z\right) \\
&= \mathbb{P}\left(\sum_{t=1}^{T}(-I_t) \leq zT - \sum_{t=1}^{T}\lambda_t^0\right) \\
&= \mathbb{P}\left(\sum_{t=1}^{T} I_t \geq -zT + \sum_{t=1}^{T}\lambda_t^0\right)
\end{aligned}
$$

where $\sum_{t=1}^{T} I_t$ is distributed as a Binomial Poisson of parameter $\{\lambda_t\}$. We observe that P_{Z_3} is an increasing function of $\{\lambda_t\}$ (i.e. P_{Z_3} shifts to left when λ_t increases). As a consequence, given a significance level α, we reject the null hypothesis when the p-value $p = P_{Z_3}(z)$ is smaller than α.

In the empirical analysis we conduct $M = 10000$ simulations using the same assumptions on the asset return distribution as for the risk measures computation. We set the test significance level α at 10%.

This test allows to verify how the choice of the asset return distribution influences the risk coverage capacity of ΛVaR, that, instead, is not directly assessed by Test 1 and Test 2. Hence, the best use of Test 3 is comparing the results between the same kind of ΛVaR models, but estimated with different assumptions on the P&L distribution (i.e. Historical, Montecarlo Normal and GARCH, etc.).

The limit of this test is that requires a massive storage of information, since at time T we need all the predictive distributions P_t of the returns for $t = 1, \ldots, T$.

4 Empirical analysis

In this section, we provide an empirical analysis of the backtesting methods of ΛVaR that we have defined in Section (3). We applied the tests to a slightly different version of the $1\% - \Lambda VaR$ models proposed in Hitaj, Mateus, and Peri (2015) and to the $1\% - VaR$ model. We compare the backtesting results with the Kupiec-type test proposed in Hitaj, Mateus, and Peri (2015) for ΛVaR and with the classical Kupiec's test for VaR.

We refer to the same dataset as in Hitaj, Mateus, and Peri (2015), consisting in daily data of 12 stocks quoted in different countries along different time windows throughout the global financial crisis (specifically, from January 2005 to December 2011). These comprise the stocks of Citigroup Inc. (C UN Equity) and Microsoft Corporation (MSFT UW Equity) for the United States, Royal Bank of Scotland Group PLC (RBS LN Equity) and Unilever PLC (ULVR LN Equity) for the United Kingdom, Volkswagen AG (VOW3 GY Equity) and Deutsche Bank AG (DBK GY Equity) for Germany, Total SA (FP FP Equity) and BNP Paribas SA (BNP FP Equity) for France, Banco Santander SA (SAN SQ Equity) and Telefonica SA (TEF SQ Equity) for Spain, and Intesa Sanpaolo SPA (ISP IM Equity) and Enel SPA (ENEL IM Equity) for Italy. The market benchmarks for the ΛVaR computation have been chosen among the market indexes with the highest volume of exchanges; these are S&P500, FTSE 100, and EURO STOXX 50.

The computation of the risk measures is based on different assumptions on the distribution of the asset returns. We consider the classical Historical and Normal simulation approach and we add robustness to the analysis by implementing GARCH models with t-student increments and the Extreme Value Theory (EVT) method based on the generalised Pareto distribution (we remand to McNeil, Frey, and Embrechts (2005) for a review on this method). The estimation of the parameters is based on 250 days of observations for the Historical and Normal assumption, while 500 days are considered for the GARCH model. For the Extreme Value Theory method, we implement an automatic routine to identify the threshold in the different time windows.

The backtesting exercise is conducted comparing the realized ex-post daily returns with the daily VaR and ΛVaR estimates of the 12 stocks over the time period of 1 year. In particular, we split the analysis into six different 2-year time windows (250 days for the risk measure computation and 1 year for the backtesting).

4.1 Results

4.1.1 Violations and Kupiec test

We first report the results of the violations and the Kupiec test for the VaR model and the Kupiec-type test adapted by Hitaj, Mateus, and Peri (2015) for the ΛVaR model. We compute the average number of violations and acceptance rate over all the assets and different time horizon T. The results presented, hereafter, in Table (1) are under the assumption of Historical distribution of the asset returns.

		Average number of violations						Kupiec-Test					
		2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011
VaR	*1%*	3.42	5.33	11.58	0.75	3.08	6.83	100 %	83 %	0 %	100 %	92 %	50 %
		3.42	**5.33**	**11.58**	**0.75**	**3.08**	**6.83**	**100%**	**83%**	**0%**	**100%**	**92%**	**50%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	2.25	3.67	7.00	0.67	2.00	4.25	100 %	83 %	42 %	100 %	100 %	83 %
	(VaR 1%)	2.17	2.33	5.75	0.67	1.58	4.00	100 %	83 %	67 %	100 %	100 %	83 %
		2.21	**3.00**	**6.38**	**0.67**	**1.79**	**4.13**	**100 %**	**83 %**	**54 %**	**100 %**	**100 %**	**83 %**
ΛVaR 1% *(incr)*	*(VaR 5%)*	1.17	1.00	3.92	0.42	0.92	2.75	100 %	100 %	100 %	100 %	100 %	100 %
	(VaR 1%)	1.17	1.08	3.92	0.42	1.00	2.75	100 %	100 %	100 %	100 %	100 %	100 %
		1.17	**1.04**	**3.92**	**0.42**	**0.96**	**2.75**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**	**100 %**

Table 1: Time evolution of the average number of violations and Kupiec test under the Historical distribution assumption. The table shows the evolution over the global financial crisis of the average number of violations and the percentage of Kupiec acceptance, aggregated at the level of $1\% VaR$, as well as the increasing and decreasing ΛVaR models.

As expected and already pointed out in Hitaj, Mateus, and Peri (2015) the average number of violations of 1% VaR is bigger than the one of ΛVaR, in particular if compared with the increasing models. In fact 1% VaR shows a drastic increase in the average number of violations, moving from 3.42 in 2006 to 11.58 in 2008. On the other hand, the increasing ΛVaR models register an average number of violations of around 1.17 during 2006 and retain the number at around 3.92 in the 2008 crisis.

This result was expected since the Λ function has been built with $\max_x \Lambda_t(x) = 0.01$, which implies that ΛVaR is always greater or equal than 1% VaR, so that, losses not covered by the first are also not covered by the latter. This implies that ΛVaR performs always better than 1% VaR by using an unilateral Kupiec-type test, since this kind of test does not capture the variability of the Λ function that is the essential feature of ΛVaR.

The violations trend is the same also under the other distribution's assumptions taken in exam as shown in Table (2).

		Normal						***GARCH***						***EVT***					
		2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011
VaR	*1%*	4.58	7.08	14.92	1.75	4.17	9.42	3.17	6.83	8.25	0.33	0.75	4.33	3.42	5.33	11.58	0.83	3.08	6.92
		4.58	**7.08**	**14.92**	**1.75**	**4.17**	**9.42**	**3.17**	**6.83**	**8.25**	**0.33**	**0.75**	**4.33**	**3.42**	**5.33**	**11.58**	**0.83**	**3.08**	**6.92**
ΛVaR 1% *(decr)*	*(VaR 5%)*	4.42	6.75	14.25	1.58	3.75	9.17	3.08	5.83	7.33	0.33	0.42	4.25	2.33	2.33	7.25	0.75	1.75	4.25
	(VaR 1%)	4.25	5.83	13.08	1.42	3.42	8.58	2.75	4.75	6.42	0.25	0.33	3.92	2.08	2.08	6.92	0.75	1.67	4.08
		4.33	**6.29**	**13.67**	**1.50**	**3.58**	**8.88**	**2.92**	**5.29**	**6.88**	**0.29**	**0.38**	**4.08**	**2.21**	**2.21**	**7.08**	**0.75**	**1.71**	**4.17**
ΛVaR 1% *(incr)*	*(VaR 5%)*	3.33	4.75	10.83	0.92	2.75	6.67	1.25	2.67	3.58	0.00	0.17	1.42	1.25	1.00	4.25	0.42	0.92	2.75
	(VaR 1%)	3.33	5.08	11.67	1.17	3.00	7.00	1.25	2.83	3.50	0.00	0.33	1.42	1.25	1.25	4.33	0.42	1.00	2.75
		3.33	**4.92**	**11.25**	**1.04**	**2.88**	**6.83**	**1.25**	**2.75**	**3.54**	**0.00**	**0.25**	**1.42**	**1.25**	**1.13**	**4.29**	**0.42**	**0.96**	**2.75**

Table 2: Time evolution of the average number of violations under the Normal, GARCH and EVT model. The table shows the evolution over the global financial crisis of the average number of violations aggregated at the level of 1%VaR, as well as the increasing and decreasing ΛVaR models.

4.1.2 Test 1 and Test 2: comparison of VaR and ΛVaR risk coverage

In Table (3) and (4) we show the results of Test 1 and 2 proposed in Section (3) for ΛVaR. The results here presented are under different assumptions of the distribution of the assets return, specifically, Historical, Normal, GARCH and EVT method.

		Historical						***Normal***					
		2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011
VaR	*1%*	100%	58%	0%	100%	75%	25%	58%	33%	0%	92%	50%	8%
		100%	**58%**	**0%**	**100%**	**75%**	**25%**	**58%**	**33%**	**0%**	**92%**	**50%**	**8%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	100 %	75 %	17 %	100 %	92 %	67 %	42%	8%	0%	83%	50%	8%
	(VaR 1%)	92 %	83 %	25 %	100 %	100 %	75 %	33%	25%	0%	92%	42%	8%
		96%	**79%**	**21%**	**100%**	**96%**	**71%**	**38%**	**17%**	**0%**	**88%**	**46%**	**8%**
ΛVaR 1% *(incr)*	*(VaR 5%)*	75 %	83 %	0 %	100 %	83 %	25 %	0 %	0 %	0 %	42 %	33 %	8 %
	(VaR 1%)	75 %	83 %	0 %	100 %	75 %	17 %	8 %	8 %	0 %	42 %	42 %	8 %
		75%	**83%**	**0%**	**100%**	**79%**	**21%**	**4%**	**4%**	**0%**	**42%**	**38%**	**8%**

		GARCH						***EVT***					
VaR	*1%*	75%	50%	33%	100%	100%	67%	100%	58%	0%	100%	75%	25%
		75%	**50%**	**33%**	**100%**	**100%**	**67%**	**100%**	**58%**	**0%**	**100%**	**75%**	**25%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	75%	50%	33%	100%	100%	67%	100 %	92 %	8 %	100 %	92 %	50 %
	(VaR 1%)	67%	67%	33%	100%	100%	67%	100 %	92%	8 %	100 %	100 %	58 %
		71%	**58%**	**33%**	**100%**	**100%**	**67%**	**100%**	**92%**	**8%**	**100%**	**96%**	**54%**
ΛVaR 1% *(incr)*	*(VaR 5%)*	67%	58%	25%	100%	92%	58%	67 %	83 %	0 %	100 %	83 %	17 %
	(VaR 1%)	75%	50%	25%	100%	92%	58%	67 %	67 %	0 %	100 %	75 %	25 %
		71%	**54%**	**25%**	**100%**	**92%**	**58%**	**67%**	**75%**	**0%**	**100%**	**79%**	**21%**

Table 3: Time evolutions of Test 1 for the ΛVaR models under different assumptions of the P&L distribution. The table shows the evolution over the global financial crisis of the acceptance rates, aggregated at the level of the ΛVaR models ($\min_x \Lambda(x) = 0.5\%$) calculated using the Historical, Normal, GARCH and EVT assumption of the P&L distribution.

		Historical						*Normal*					
		2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011
VaR	*1%*	100 %	75 %	0 %	100 %	92 %	42 %	58%	42%	0%	100%	67%	25%
		100%	**75%**	**0%**	**100%**	**92%**	**42%**	**58%**	**42%**	**0%**	**100%**	**67%**	**25%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	100%	83%	17%	100%	100%	75%	58%	42%	0%	100%	67%	17%
	(VaR 1%)	100%	83%	42%	100%	100%	83%	50%	50%	0%	100%	67%	17%
		100%	**83%**	**29%**	**100%**	**100%**	**79%**	**54%**	**46%**	**0%**	**100%**	**67%**	**17%**
ΛVaR 1% *(incr)*	*(VaR 5%)*	100%	100%	17%	100%	92%	42%	17%	25%	0%	92%	50%	8%
	(VaR 1%)	100%	100%	17%	100%	92%	42%	25%	33%	0%	83%	58%	25%
		100%	**100%**	**17%**	**100%**	**92%**	**42%**	**21%**	**29%**	**0%**	**88%**	**54%**	**17%**
		GARCH						*EVT*					
VaR	*1%*	83%	58%	42%	100%	100%	67%	100 %	75 %	0 %	100 %	92 %	33 %
		83%	**58%**	**42%**	**100%**	**100%**	**67%**	**100%**	**75%**	**0%**	**100%**	**92%**	**33%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	83%	58%	33%	100%	100%	67%	100 %	92 %	8 %	100 %	100 %	67 %
	(VaR 1%)	92%	75%	42%	100%	100%	75%	100 %	92 %	17 %	100 %	100 %	67 %
		88%	**67%**	**38%**	**100%**	**100%**	**71%**	**100%**	**92%**	**13%**	**100%**	**100%**	**67%**
ΛVaR 1% *(incr)*	*(VaR 5%)*	92%	75%	67%	100%	100%	83%	100 %	100 %	17 %	100 %	92 %	42 %
	(VaR 1%)	92%	67%	67%	100%	92%	83%	100 %	92 %	17 %	100 %	92 %	42 %
		92%	**71%**	**67%**	**100%**	**96%**	**83%**	**100%**	**96%**	**17%**	**100%**	**92%**	**42%**

Table 4: Time evolutions of Test 2 for the ΛVaR models under different assumptions of the P&L distribution. The table shows the evolution over the global financial crisis of the acceptance rates, aggregated at the level of the ΛVaR models ($\min_x \Lambda(x) = 0.5\%$) calculated using the Historical, Normal, GARCH and EVT assumption of the P&L distribution.

We first notice that the acceptance rate of these tests is lower than the unilateral Kupiec test in Hitaj, Mateus, and Peri (2015). This is due to the particular construction of the Kupiec test. Indeed, this test is useful to assess if the ΛVaR model guarantees an acceptable coverage given by $\max(\Lambda)$, but cannot capture the daily variations of the confidence level λ_t^0 of ΛVaR. Thus, it cannot be used to evaluate the real coverage offered by ΛVaR at time t. On the other hand, the coverage tests that we have proposed are able to better evaluate if the flexibility introduced by the Λ function helps to detect adverse scenario and put aside a more adequate amount of capital.

If we compare the tests results, we observe that for all the models Test 2 provides higher acceptance rates in respect to Test 1. This may be due to the fact that Test 1 returns more precise results with smaller number of observations and also to its unilateral nature that attributes the highest weight to the violations.

With the exception of the normal estimator, the ΛVaR models result often more accurate than 1% VaR, confirming the outcomes in Hitaj, Mateus, and Peri (2015). This means that the highest flexibility of ΛVaR contributes to the highest coverage, especially when it is computed with distributions that better capture the tail behaviour. In our tests, the decreasing ΛVaR models seem to be more accurate, in contrast with the results of the Kupiec test. We think this is a consequence of a lower power of these tests for the decreasing ΛVaR models. We remand the analysis of the test power for further research since it would complicate this study without adding significant value.

4.1.3 The choice of the Λ minimum

During the analysis of the results, Test 1 and Test 2 have pointed out an issue of estimation in the ΛVaR models proposed by Hitaj, Mateus, and Peri (2015). In particular, the authors do not discuss in details the choice of the Λ minimum, $\min_x \Lambda(x)$, that seems to be set equal to 0.1% after empirical experimentations. In addition, the extended Kupiec test proposed by the authors could not identify the impact of this choice.

When we have run for the first time Test 1 and 2 using the choice of Hitaj, Mateus, and Peri (2015), $\min_x \Lambda(x) = 0.1\%$, we have noticed that the increasing ΛVaR models presented the highest rejection rate, even if they had the smallest number of infractions, as shown by Table (5).

		Test 1						***Test 2***					
		2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011
ΛVaR 1% *(decr)*	*(VaR 5%)*	100 %	75 %	8 %	100 %	92 %	67 %	100 %	83 %	17 %	100 %	100 %	75 %
	(VaR 1%)	92 %	83 %	25 %	100 %	100 %	67 %	100 %	83 %	42 %	100 %	100 %	83 %
		96 %	**79 %**	**17 %**	**100 %**	**96 %**	**67 %**	**100 %**	**83 %**	**29 %**	**100 %**	**100 %**	**79 %**
ΛVaR 1% *(incr)*	*(VaR 5%)*	8 %	17 %	0 %	58 %	42 %	8 %	75 %	83 %	0 %	100 %	75 %	17 %
	(VaR 1%)	8 %	17 %	0 %	58 %	42 %	8 %	75 %	83 %	0 %	100 %	83 %	25 %
		8 %	**17 %**	**0 %**	**58 %**	**42 %**	**8 %**	**75 %**	**83 %**	**0 %**	**100 %**	**79 %**	**21 %**

Table 5: Time evolutions of Test 1 and Test 2 for the ΛVaR models with $\min_x \Lambda(x) = 0.1\%$ under the Historical distribution assumption. The table shows the evolution over the global financial crisis of the acceptance rates, aggregated at the level of the ΛVaR models with $\min_x \Lambda(x) = 0.1\%$.

Thus, we have studied how the probability of infraction λ_t evolves in the different ΛVaR models and we have observed that in most of the cases it obtains the minimal value. This happens especially during crisis periods, when the cumulative distribution function of the assets shifts on the left and intersects the Λ function at the minimum level. In such a case, the choice of the Λ minimum is relevant and also a critical issue.

From our point of view, the Λ minimum should provide the probability to lose more than the *worst case event* (i.e. benchmarks' minimum, $\pi_1 = \min x_{t,j}$) over the time window observations (i.e. 250 in our case). If we consider all the events equally probable, the selection of the Λ minimum should be greater than $1/T$ over T observations. Thus, we propose to compute the ΛVaR models by fixing the Λ minimum equal to 0.5%, i.e. $\min_x \Lambda(x) = 0.005$, since the probability of an event over 250 past realizations is 0.4%.

The results of the ΛVaR estimations with 0.5% minimum have been shown before in

Table (3) and (4). The number of infractions does not change in any period under consideration, while the acceptance rate of the increasing ΛVaR models drastically increases, validating our choice. Clearly, this new setting does not affect the decreasing ΛVaR models. Anyway, the choice of the Λ minimum can be refined considering more precise evaluation of the probability of the worst case event, but this is beyond the objective of this paper.

4.1.4 Test 3: comparison of $\Lambda VaRs$ with different distribution estimations

As anticipated in Section (3), the best use of Test 3 is the comparison of the accuracy of the risk measures computed with different estimations of asset return distribution. We compute the time evolution of the acceptance rate aggregated at the level of the increasing and decreasing ΛVaR models. We repeat the analysis changing the assumption on the asset return distribution: specifically, Historical, Monte Carlo Normal, GARCH and EVT method. The results are presented in Table (6)

		Historical						***Normal***					
		2006	2007	2008	2009	2010	2011	2006	2007	2008	2009	2010	2011
VaR	*1%*	50%	33%	0%	100%	58%	25%	58%	33%	0%	92%	50%	8%
		50%	**33%**	**0%**	**100%**	**58%**	**25%**	**58%**	**33%**	**0%**	**92%**	**50%**	**8%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	50%	33%	0%	100%	67%	17%	58%	42%	0%	92%	58%	17%
	(VaR 1%)	58%	50%	8%	100%	67%	8%	50%	33%	0%	92%	58%	25%
		54%	**42%**	**4%**	**100%**	**67%**	**13%**	**54%**	**38%**	**0%**	**92%**	**58%**	**21%**
ΛVaR 1% *(incr)*	*(VaR 5%)*	8%	17%	0%	58%	42%	0%	17%	17%	0%	92%	50%	8%
	(VaR 1%)	8%	17%	0%	58%	42%	8%	33%	8%	0%	83%	50%	17%
		8%	**17%**	**0%**	**58%**	**42%**	**4%**	**25%**	**13%**	**0%**	**88%**	**50%**	**13%**
		GARCH						***EVT***					
VaR	*1%*	75%	58%	33%	100%	100%	67%	50 %	33 %	0 %	100 %	58 %	25 %
		75%	**58%**	**33%**	**100%**	**100%**	**67%**	**50%**	**33%**	**0%**	**100%**	**58%**	**25%**
ΛVaR 1% *(decr)*	*(VaR 5%)*	75%	58%	33%	100%	100%	67%	67 %	58 %	0 %	92 %	58 %	42 %
	(VaR 1%)	92%	67%	33%	100%	100%	75%	67 %	58 %	0 %	92 %	58 %	33 %
		83%	**63%**	**33%**	**100%**	**100%**	**71%**	**67%**	**58%**	**0%**	**92%**	**58%**	**38%**
ΛVaR 1% *(incr)*	*(VaR 5%)*	83%	67%	67%	100%	100%	83%	17 %	25 %	8 %	58 %	42 %	0 %
	(VaR 1%)	83%	58%	67%	100%	92%	83%	17 %	33 %	0 %	58 %	42 %	8 %
		83%	**63%**	**67%**	**100%**	**96%**	**83%**	**17%**	**29%**	**4%**	**58%**	**42%**	**4%**

Table 6: Time evolutions of Test 3 for the ΛVaR models under different assumptions of the P&L distribution. The table shows the evolution over the global financial crisis of the acceptance rates, aggregated at the level of the ΛVaR models ($\min_x \Lambda(x) = 0.5\%$) calculated using the Historical, Normal, GARCH and EVT assumption of the P&L distribution.

The results show that the GARCH assumption on the returns guarantees the highest accuracy in terms of average acceptance rate. Moreover, we notice here that the Historical and the EVT estimators of the increasing ΛVaR often underperform the Normal one, in contrast with the previous tests. These outcomes are quite reasonable since this third test is based on simulations and points out the issue of estimating risk measures with distributions having cut-off tails (as the Historical) or based on a small range of values (as the EVT). However, such a preference for the Normal distribution is completely reversed by the other tests which privilege the assumption of distributions which rely more on tail events and not on the full shape of the distribution.

5 Conclusions

A new risk measure sensitive to tail risk, ΛVaR, has been recently introduced. However, an *ad hoc* study on its backtesting has not been conducted in literature so far. The main issue for the ΛVaR backtesting is that the probability of a violation is not constant, but may change at any time and for any asset. This consideration implies that the Kupiec-type backtesting framework, proposed by Hitaj, Mateus, and Peri (2015), fails to keep into account the effective predictive capacity of ΛVaR as introduced by the Λ function.

We propose three backtesting methodologies for ΛVaR and we asses the accuracy of the new risk measure from different points of view. Test 1 and Test 2 are based on results of probability theory and allow for a straightforward application. Test 3 is performed by simulations and allows for more accurate comparison of ΛVaR models estimated under different assumptions on the P&L distribution.

The validity of our backtesting proposals is confirmed by the results of the empirical analysis. In fact, this study shows that ΛVaR models perform better than 1 % VaR, confirming the findings in Hitaj, Mateus, and Peri (2015). In addition, ΛVaR computed with the GARCH model of returns has the highest level of coverage. This outcome substantiates what is well known in literature that fat-tailed asset return distributions explain better the real asset return behavior and allow for a more accurate risk coverage.

Moreover, our backtesting methods denote higher precision than the Kupiec-type test proposed by Hitaj, Mateus, and Peri (2015) since they have been able to detect an estimation issue of ΛVaR computed with a lower bound of 0.1% as in the former study.

Suggestions for future research include the study of the test power that would permit a more accurate comparison among these backtesting proposals.

Appendix

We recall hereafter the Lyapunov Theorem that is a result of probability theory based on the application of the central limit theorem to random variables that are independent but not identically distributed (see Lyapunov 1954).

Theorem 1 (Lyapunov) *Suppose $X_1, X_2, ...$ is a sequence of independent random variables, each with finite expected value μ_t and variance σ_t^2. Define*

$$s_n^2 = \sum_{t=1}^{T} \sigma_t^2$$

If for some $\delta > 0$, the "Lyapunov's condition"

$$\lim_{n\to\infty} \frac{1}{s_T^{2+\delta}} \sum_{t=1}^{T} \mathrm{E}\big[\, |X_t - \mu_t|^{2+\delta} \,\big] = 0$$

is satisfied, then the following convergence in distribution holds as T goes to infinity:

$$\frac{1}{s_T} \sum_{t=1}^{T} (X_t - \mu_t) \ \xrightarrow{d}\ \mathcal{N}(0,\ 1)$$

In the following lemma we show that the "Lyapunov's condition" is satisfied when $s_T^2 = \sum_1^T \lambda_t(1-\lambda_t)$ and $\mu_t = \lambda_t$.

Lemma 2 *If $\{I_t\}$ is a sequence of independent random variables distributed as a Bernoulli with parameters $\{\lambda_t\}_t$ and $\inf_t \lambda_t = \lambda_m > 0$, then*

$$\lim_{T\to\infty} \frac{1}{s_T^{2+\delta}} \sum_{t=1}^{T} \mathrm{E}[|I_t - \lambda_t|^{2+\delta}] = 0$$

with $s_T^2 = \sum_1^T \lambda_t(1-\lambda_t)$.

Proof. We observe that:

$$\begin{aligned}\mathrm{E}[|I_t - \lambda_t|^{2+\delta}] &= (1-\lambda_t)\lambda_t^{2+\delta} + \lambda_t(1-\lambda_t)^{2+\delta} \\ &= \lambda_t(1-\lambda_t)\left(\lambda_t^{1+\delta} + (1-\lambda_t)^{1+\delta}\right) \le \lambda_t(1-\lambda_t) \le \frac{1}{4}\,.\end{aligned}$$

On the other hand we have

$$s_T^{2+\delta} = \left(\sum_1^T \lambda_t(1-\lambda_t)\right)^{1+\frac{\delta}{2}} \ge \left(\sum_1^T \lambda_m(1-\lambda_m)\right)^{1+\frac{\delta}{2}} = (T\lambda_m(1-\lambda_m))^{1+\frac{\delta}{2}}\ .$$

We can thus conclude that

$$\frac{\sum_{t=1}^T \mathrm{E}[|I_t-\lambda_t|^{2+\delta}]}{s_T^{2+\delta}} \le \frac{T}{4\,(T\lambda_m(1-\lambda_m))^{1+\frac{\delta}{2}}} \to 0$$

as $T \to \infty$. ■

The following proposition shows theoretical implications on the Z_3 test statistic in (18) under the null and alternative hypothesis.

Proposition 3 *Under the test hypothesis H_0 as in (9) and H_1 as in (11) we have:*

1. $\mathrm{E}_{H_0}[Z_3] = 0$

2. $\mathrm{E}_{H_1}[Z_3] < 0$.

Proof. It is enough to notice that under H_0, $I_t \sim B(\lambda_t^0)$ so that $\mathrm{E}_{H_0}[I_t - \lambda_t^0] = 0$, which implies

$$\mathrm{E}_{H_0}[Z_3] = \frac{1}{T} \sum \mathrm{E}_{H_0}[\lambda_t^0 - I_t] = 0\,.$$

In a similar way, under H_1, since $I_t \sim B(\lambda_t)$ with $\lambda_t > \lambda_t^0$, we obtain that $\mathrm{E}_{H_1}[Z_3] < 0$.

■

References

Acerbi, C., and B. Szekely. 2014. "Back-testing Expected Shortfall." Risk 27 (11): 76-81.

Berkowitz, J., P. Christoffersen, and D. Pelletier. 2011. "Evaluating Value-at-Risk Models with Desk-Level Data." Management Science 57: 2213-2227.

Basel Committee on Banking Supervision. 1996. "Supervisory Framework for the Use of Backtesting in Conjunction with the Internal Models Approach to Market Risk Capital Requirements." Bank for International Settlements.

Basel Committee on Banking Supervision. 2013. "Fundamental Review of the Trading Book." 2nd consultative document. Bank for International Settlements.

Campbell, S. 2005. "A Review of Backtesting and Backtesting Procedures." Finance and

Economics Discussion Series. Divisions of Research & Statistics and Monetary Affairs, Federal Reserve Board, Washington D.C.

Christoffersen, P. 2010. "Encyclopedia of Quantitative Finance - Backtesting." Wiley Online Library.

Frittelli, M., M. Maggis, and I. Peri. 2014. "Risk Measures on and Value at Risk with Probability/Loss Function." Mathematical Finance 24 (3): 442-463.

Hitaj, A., C. Mateus, and I. Peri. 2015. "Lambda Value at Risk and Regulatory Capital: a Dynamic Approach to Tail Risk." Working paper available at http://ssrn.com/abstract=2932475.

Kupiec, P. 1995. "Techniques for Verifying the Accuracy of Risk Measurement Models." Journal of Derivatives 3 (2): 73-84.

Lyapunov, A. M. 1954. "Collected Works 1." Moscow-Leningrad, 157-176.

McNeil, A. J., R. Frey, and P. Embrechts. 2005. "Quantitative Risk Management: Concepts, Techniques and Tools." Princeton University Press, 3 Market Place, Woodstock, Oxfordshire OX20 1SY, United Kingdom.