IJEFI

INTERNATIONAL JOURNAL OF ECONOMICS AND FINANCIAL ISSUES

EJ EconJournals

ISSN: 2146-4138

International Journal of Economics and Financial Issues

ISSN: 2146-4138

available at http: www.econjournals.com

EJ EconJournals

International Journal of Economics and Financial Issues, 2022, 12(3), 38-52.

On the Effectiveness of Stock Index Futures for Tail Risk Protection

Hammadi Zouari*

Department of Accounting and Finance, High Institute of Management, Gabes, Tunisia. *Email: zouari_hamadi@yahoo.fr

Received: 15 February 2022 **Accepted:** 16 April 2022 **DOI:** https://doi.org/10.32479/ijefi.13011

ABSTRACT

This paper examines the effectiveness of using stock index futures contracts as substitutes for fixed-income securities in implementing expected shortfall targeting strategy. We find that the futures-based implementation outperforms its index-and-bill counterpart both in terms of downside protection and risk-adjusted performance at daily rebalancing frequency. This outperformance is driven not only by the transaction cost advantage, but also by the replication imperfections due to futures mispricing providing over the long term a better participation in upward market movements. When less frequent rebalancing intervals are used, the futures-based implementation becomes less effective at protecting the downside risk but still capture better the upside potential of the index.

Keywords: Stock Index Futures, Tail Risk Protection, Target Risk Strategies, Value-at-Risk, Expected Shortfall, Extreme Value Theory
JEL Classifications: C53, G11

1. INTRODUCTION

Recurring financial crises have contributed to an extensive strand of literature focused on tail risk protection strategies designed to mitigate tail risk in order to protect investment portfolios against the dire effects of a significant market down draught.[1] The growing interest for tail risk and its mitigation is largely motivated by the serious limitations of portfolio diversification as a risk management tool. This is primarily owed to the empirical observation that correlations across asset classes tend to increase during market drawdowns and become more intense in periods of crisis. For example, Chabi-Yo et al. (2018) find an increased dependence in the left tail of stock returns in times of market crashes.[2] Additionally to the simultaneous increase of correlations, Bollerslev et al. (2018) observe strong similarities in realized volatilities patterns within and across equities, bonds, commodities and currencies. Furthermore, Jondeau and Rockinger (2003) find that also co-movements of higher moments of stock-index and foreign-exchange returns are strongly related and become more intensive during agitated periods. According to Bhansali (2011) and Benson et al. (2013), the reason behind all these dependencies is that many non-equity asset classes are exposed to an equity market risk factor that explains the largest portion of cross-sectional asset class return variance. Additionally, other relevant factors such as increased globalization, stock market contagion and liquidity shocks tend to reduce portfolio diversification benefits (Poon et al., 2004).

Given these limitations, the financial research has been oriented towards alternative investments strategies likely to involve substantial reduction in a portfolio's downside risk while simultaneously keeping most of its upside potential. Among these, the risk targeting strategies have received extensive attention

1 Tail risk refers to the extreme losses at the left tail of an asset's or portfolio's return distribution return.

2 Longin and Solnik (2001), Campbell et al. (2002), Buraschi et al. (2010) and Christoffersen et al. (2012) among others for studies with a focus on time-varying asset class correlations.

from both academics and practitioners.[3] They aim to keep the *ex-ante* risk of a portfolio, typically consisting of a risky asset and a riskless asset, at a pre-specified constant target level. To achieve this, the investment exposure of the portfolio is continually adjusted in inverse relation to updated risk forecasts of the risky asset, i.e. upwards if the predicted risk decreases and downwards if it increases. By doing so, the risk targeting strategies control the portfolio risk over time by taking advantage of the so-called leverage effect whereby, at least for equities and indices, the returns tend to be negatively correlated with risk, which is in turn allows capturing the upside potential of the risky asset while simultaneously drawdowns are mitigated (Wang et al., 2012; Hallerbach, 2015; Harvey et al., 2018).

The majority of prior research has concentrated on target volatility strategy. Giese (2012) demonstrated that it always delivers a better Sharpe ratio than the underlying index in the long term as long as the volatility target is chosen below a certain threshold, irrespective of the underlying index and the volatility distribution. He also argued that the improvement in the Sharpe ratio is positively related to the volatility of volatility because more variable equity volatility creates more opportunities for the strategy to outperform the underlying index by responding to volatility. Hallerbach (2012, 2015) proved a similar finding that the volatility targeting mechanism improves the risk-adjusted performance even if the portfolio mean return is constant over time and outlines two theoretical arguments for the superiority of managed volatility portfolios: the negative relationship between the return and volatility and the performance enhancing abilities of risk control over time.

Risk targeting strategies have been extensively backtested using historical data (Cooper, 2010; Giese, 2012; Ilmanen and Kizer, 2012; Kirby and Ostdiek, 2012; Hocquard et al., 2013). All these papers provide evidence that the application of a volatility targeting strategy for equities offers an enhanced risk-return performance compared to a buy-and-hold equity investment, despite a lower return due to the drag on performance during bull markets. Dreyer and Hubrich (2019) find that the Sharpe ratio benefit of the application of volatility targeting strategy for equities is period dependent. Comparing across asset classes, Perchet et al. (2016) and Harvey et al. (2018) observe that the volatility targeting reduces tail risks irrespective of the asset class, which is the main driver of the outperformance of volatility targeting, while the improvement in the Sharpe ratio is limited to equities and corporate bonds due to the leverage effect for those assets. Dachraoui (2018) theoretically shows that a negative risk-return relation is not needed to provide an enhanced risk-return profile of volatility targeting. Furthermore, Bollerslev et al. (2018) show that using realized volatility models in a volatility targeting strategy translated into better risk-adjusted performance and higher utility gains in comparison with less accurate forecasting models that do not incorporate the information in high-frequency intraday data. A similar result is also found by Perchet et al. (2016), Moreira and Muir (2017) and di Persio et al. (2021).

3 Other strategies include mainly the methods of portfolio insurance such as the option based portfolio insurance (OBPI) and the constant proportion portfolio insurance (CPPI) as well as its dynamic extension (DPPI).

As stated above, previous studies on risk targeting have almost exclusively used volatility as a measure of risk. However, for tail risk protection, a downside risk measure should be preferred to volatility for at least three reasons. First, prospect theory asserts that most investors seek upside potential and downside protection, i.e. they are risk-seeking above a certain threshold return, but risk-averse below this threshold. Loss-averse investors should therefore adopt strategies that deal with negative returns and not return deviations to enhance their utility (Ang et al., 2006a; Bollerslev et al., 2015). Second, due to the empirical facts that asset returns are typically asymmetric and exhibit excess kurtosis, volatility underestimates the extreme events of low probability and can thus lead to insufficient protection against extreme negative market moves (Poon et al., 2004, Gormsen and Jensen, 2020) and/or severe drag on performance in times of huge positive returns. Third, downside risk measures implicitly incorporate the information contained in higher moments, which may enhance the effectiveness of the tail risk protection strategies in particular in periods of high market volatility when the return distributions tend to be more negatively skewed and fait-tailed (Bali et al., 2009; Gormsen and Jensen, 2020). Consequently, target a set level of downside risk, measured by value-at-risk (VaR) or expected shortfall (ES), instead of volatility is better suited for loss control in severe market downturns. This is confirmed by Strub (2013) and Rickenberg (2020) who find that tail risk targeting strategies outperform the volatility based strategy in terms of risk-adjusted performance and drawdown in times of bear markets. Rickenberg (2020) also observes that the performance can be enhanced by switching between volatility and ES targeting based on estimates of whether the market will be in a bull or bear regime. Happersberger et al. (2020) conduct a historical block-bootstrap analysis for the ES targeting strategy applied to a multi-asset portfolio and find that this strategy is more profitable when the portfolio ES is estimated using a forecast combination technique based on a loss function of Fissler and Ziegel (2016).

Risk targeting strategies are typically implemented by combining an equity index fund with a money market instrument like treasury bills and then shifting money between them to keep the risk of the portfolio constant over time. This implementation strategy requires frequent trading in the index which can lead to high transaction costs. A possible alternative, that would achieve the same goal without ever trading in the index, is to short exchange-listed futures written on the index (Hocquard et al., 2013; Papageorgiou et al, 2017; Bongaerts et al., 2020). The short position is then adjusted dynamically to changes in a risk forecast, i.e. shorting more futures contracts when the risk forecast increases and reduce the short position as the risk forecast falls.

The advantage of using stock index futures is that they are highly liquid and transaction costs are small. This raises the question of whether the cost advantage of the futures-based implementation translates into better risk-adjusted performance and downside protection. This question has previously been addressed for dynamic insurance portfolio strategies. Using futures contracts on the Australian all ordinaries index, Do and Faff (2004) performed historical simulations for the synthetic put and CPPI strategies. Based on daily rebalancing, they find that the futures-based implementation achieves the desired floor more often and incurs a lower cost of insurance than its index-and-bill rival during the period from March 1992 to December 2002, using either the synthetic put or CPPI approach. By regenerating the simulations based on theoretical futures prices from the cost-of-carry model, it turned out that the underperformance of the futures-based implementation during the period from March 1988 to December 1991 is owed to the futures mispricing problem. Furthermore, when less frequent rebalancing intervals are used, for 1% and 2% market move triggers, the CPPI appears unaffected while the futures-based synthetic put records higher mean returns. In the same vein, Loria et al. (1991) simulated the performance of a futures-based synthetic put strategy and reported that there is no perfect guarantee of loss prevention under any scenario. They state that the strategy provides downside protection against severe market declines, but not for small downward movements.

The purpose of this paper is to provide detailed evaluations and comparisons of two procedures of implementing an ES targeting strategy: via index and bills, and via futures markets. We empirically examine the ability of the futures-based implementation to provide better downside protection and risk-adjusted performance than its index-and-bill counterpart. We seek to contribute to the existing literature by conducting an empirical comparison of the two implementation strategies in the tail risk protection framework. To the best of our knowledge, no prior studies have examined the performance of futures-based ES targeting strategies.

The remainder of this paper is structured as follows. In Section 2 we present the risk targeting strategy and describe its two implementation procedures. Section 3 briefly describes the methods employed for forecasting and backtesting VaR and ES. Section 4 presents the data and the empirical results of our study and section 5 concludes.

2. RISK TARGETING STRATEGIES

Risk targeting is a strategy that rebalances between a risky asset and a non-risky asset in order to target a constant level of risk over time. For this purpose, the portfolio's exposure is dynamically adjusted conditional on a forecast of the risky asset's risk. Specifically, the protection is performed by falling back on the risk-free asset when the portfolio's current risk is higher than the predefined target level. In the opposite scenario, the investment in the risk-free asset is reduced in favour of a higher exposure to the risky asset. The basic idea of maintaining a constant level of risk over time is to adjust market exposure inversely proportional to market risk, thereby taking advantage of the negative relationship between risk and future returns. Times of high tail risk marked by a negative skewness and/or high kurtosis are generally followed by low returns on one hand (Gormsen and Jensen, 2020), and empirical evidence suggests that bull market periods are accompanied by high returns and low risk, on the other. Therefore, by conditioning their exposure market risk, investors can therefore capture the upside potential while simultaneously drawdowns are mitigated.

Throughout the paper, we consider a risk-averse investor who seeks to protect his investment in a risky asset, e.g. an equity index, against extreme market losses over an investment period through a risk targeting strategy. The price of the risky asset at time t is denoted as S_t such that the logarithmic return over the time period from t to t+1, representing 1 day, is $r_{t+1}=\log(S_{t+1}/S_t)$. At the beginning of the investment period, the investor sets a target risk level that should be kept constant over time. To this end, one methodology is to manage a protected portfolio combining the risky asset with a risk-free asset, e.g. treasury bills. Specifically, the allocation to the risky asset has to be chosen as:

$$w_t = \frac{\bar{\rho}}{\hat{\rho}_t\left(r_{t+1}\middle|\mathcal{F}_t\right)} \tag{1}$$

where $\bar{\rho}$ represents the chosen risk target level and $\hat{\rho}_t\left(r_{t+1}\middle|\mathcal{F}_t\right)$ denotes the downside risk forecast of the risky asset from t to t+1 conditional on the available information, $\mathcal{F}_t$. Denote by V_t the value of the protected portfolio at time t, the investment exposure to the risky asset is $E_t=w_tV_t$ and the remaining funds $B_t=(1-w_t)V_t$ are invested in the riskless asset.

The protected portfolio is thus routinely adjusted based on updated risk forecasts such that the *ex ante* risk remains close to the chosen risk target level over time. In this way, the risk targeting strategy represents a consistent active approach that dynamically allocates funds between the risky and the riskless asset in response to the prevailing market risk conditions. However, with such an implementation procedure, the strategy is subject to potential liquidity constraints that prevent portfolio rebalancing and high transaction costs that can extinguish its efficiency and performance.

An alternative implementation of the risk targeting strategy, that can overcome such limitations, is to short exchange-traded futures contracts written on the underlying risky asset as monetary market transaction substitutes. In that case, the funds are fully invested in the risky asset and an overlay of long and short futures contracts is used to keep the portfolio risk in line with the risk target level. Let $r_{f,t}$ and q_t denote, respectively, the riskless asset return and the dividend yield on the risky asset at time t (both are annualized and expressed with continuous compounding). The futures-based implementation involves maintaining a short position in a number of futures:

$$n_t = \frac{B_t}{S_t e^{\left(r_{f,t}-q_t\right)(T-t)-\frac{r_{f,t}}{365}}} \tag{2}$$

Where T is the remaining life of the futures contract, expressed in years (Appendix for proof).

This equation must however be modified to avoid trading in the risky asset resulting from having to settle gains and losses in the futures market:

$$n_t = \frac{B_t - \pi_t}{S_t e^{\left(r_{f,t}-q_t\right)(T-t)-\frac{r_{f,t}}{365}}} \tag{3}$$

where π_t denotes the accumulated gains at time t resulting from futures settlement (Do and Faff, 2004).

In order to implement a risk targeting strategy, we need first to choose the risk measure to target according to performance and protection objectives. As the investor's primary objective was to mitigate extreme negative returns, targeting a tail-related risk measure, such as VaR or ES, is more appropriate than volatility. Then we must specify an adequate risk model to forecast the daily downside risk of the risky asset. The choice of this model is of paramount importance to optimize the investor's benefit from risk targeting strategy (Rickenberg, 2020; Bollerslev et al., 2018; Happersberger et al., 2020).

3. CONSTRUCTING AND EVALUATING RISK FORECASTS

In this section, we provide a short description of the two downside risk measures VaR and ES, then we briefly present the risk forecasting models considered in this paper and the main backtesting procedures employed to assess the accuracy of the forecasts.[4]

3.1. Downside Risk Measures

VaR and ES are basically the two most popular and widely used measures of downside risk (Jorion, 2021). VaR is defined as the minimum potential return of an asset (or a portfolio) over a certain time horizon for a given confidence $(1–p)$, where $p\in(0,1)$ is the significance level (typically 1% or 5%). Therefore, the p-level VaR forecast from t to t+1 corresponds to the p-quantile of the conditional return distribution function at time t+1, that is:

$$VaR_{t+1|t}^{p} = F_p^{-1}\left(r_{t+1}\middle|\mathcal{F}_t\right) = \inf\left\{r_{t+1} : F\left(r_{t+1}\middle|\mathcal{F}_t\right) \geq p\right\} \quad (4)$$

where $F(.)$ denotes the cumulative distribution function, assumed to be continuous and strictly increasing with finite mean such that its inverse function is well defined.

VaR has the main disadvantage of disregarding the risk of extreme losses beyond the confidence level which may induce large losses (Basak and Shapiro, 2001). Moreover, it does not satisfy the subadditivity, which is one of the essential proprieties of coherent risk measure, and hence does not reward diversification (Artzner et al., 1999). To cope with these shortcomings, (Artzner et al., 1999) and Basak and Shapiro (2001) propose the use of the ES as an alternative measure of risk. ES, also referred to as conditional VaR or expected tail loss, is defined as the conditional expectation of the return given that it is less than the VaR (Yamai and Yoshiba, 2002; Taylor, 2008). Specifically, the p-level ES forecast from t to t+1 can be written as:

$$ES_{t+1|t}^{p} = E\left(r_{t+1}\middle|r_{t+1} \leq VaR_{t+1}^{p}\right) = \frac{1}{p}\int_0^p VaR_{t+1}^{p}(u)\,du \quad (5)$$

ES is a coherent measure because it satisfies the subadditivity propriety and consequently can be reduced by diversification. Further, it directly controls the risk in the left tail of the return distribution, so that extreme losses beyond the confidence level are explicitly taken into account as a conditional expectation.

3.2. Risk Forecasting Models

A large number of methods for estimating VaR and ES have been proposed in the literature and are often categorized into three main categories: parametric methods, nonparametric methods and semi-parametric methods (Engle and Manganelli, 2004). Nonparametric approach has the advantage of practical and easy implementation since it requires no distributional assumptions. In contrast, parametric approach simply assume that the observed returns follow a specific probability distribution, such as a Normal or a Student's t. Semi-parametric estimation methods combine the parametric and nonparametric approaches.

In what follows, we will only focus on the most methods that are widely used by practitioners and in the academic literature.[5] In particular, all the methods presented here will be used to make one-step-ahead VaR and ES forecasts at a probability level p with a rolling window of length n.

3.2.1. Historical simulation

The simplest existing method would be the historical simulation which is one of the non-parametric approaches. If the fundamental pre-requisite for it to work is to assume that the past returns are *i.i.d.* and will recur in future (Dowd, 2005), it main difficulty lies in choosing a relevant window's width.[6] Let $r_{1,n} \leq r_{2,n} \leq \ldots \leq r_{n,n}$ denote the returns within this window (i.e. from $t+1–n$ to t) sorted in ascending order. Then, the VaR for t+1 is simply given by the $[np]$-th order statistic, that is:

$$VaR_{t+1|t}^{p} = r_{[np],n} \quad (6)$$

Correspondingly, the ES estimate for t+1 can be computed based on the ordered returns by:

$$ES_{t+1|t}^{p} = \left(\sum_{i=1}^{[np]-1} r_{i,n} + r_{[np],n}\left(np - [np] + 1\right)\right)(np)^{-1} \quad (7)$$

3.2.2. Gaussian normal distribution

Despite the typical features of financial assets such as fat tails and non-normality, the central limit theorem states that, when the sample is large enough, the normal distribution can be regarded as sufficient to fit the returns well. Assuming that the returns are normally distributed, the one-step-ahead VaR is given by:

$$VaR_{t+1|t}^{p} = \breve{\mu}_{t+1} + \breve{\sigma}_{t+1}\Phi^{-1}(p) \quad (8)$$

4 While we abstract in our empirical study from the use of VaR in the implementation of the risk targeting strategy, this section also deals with both VaR and ES since the ES tests of McNeil and Frey (2000) and Nolde and Ziegel (2017) require both VaR and ES forecasts as input variables.

5 We recommend Kuester et al. (2006), Christoffersen (2016), Andersen et al. (2011), Righi and Ceretta (2015) and Lazar and Zhang (2019) for a more comprehensive discussion of different VaR and ES estimation techniques.

6 Several implementations can be added to the basic historical simulation for improving the estimated risk measures, such as bootstrapping, weighting (according to age, volatility or correlation) and combination of nonparametric density function. Dowd (2005) chapter 4 for a rigorous discussion of different generalizations of the basic historical simulation.

where $\hat{\mu}_{t+1}$ and $\breve{\sigma}_{t+1}$ are computed by the sample mean and sample standard deviation, and Φ denotes the standard normal cumulative distribution function.

Correspondingly, the one-step-ahead ES can be derived as:

$$ES_{t+1|t}^{p} = \breve{\mu}_{t+1} - \breve{\sigma}_{t+1} \frac{\varphi\left(\Phi^{-1}(p)\right)}{p} \tag{9}$$

Where Φ denotes the probability density function of the standard Gaussian distribution.

3.2.3. Student's t distribution

The symmetric t-Student's is used to address the issue of heavy tails and more peak in the distribution of the standardized returns in comparison with the normal. If we consider that the standardized returns follow a Student's t distribution, the VaR forecast for t+1 can be computed as:

$$VaR_{t+1|t}^{p} = \breve{\mu}_{t+1} + \breve{\sigma}_{t+1}\sqrt{v^{-1}(v-2)}t_v^{-1}(p) \tag{10}$$

where $t_v^{-1}(p)$ is the p-quantile of the standardized returns following a Student's t distribution with estimated degree of freedom v larger than 2.

Correspondingly, the ES forecast for t+1 can be calculated as:

$$ES_{t+1|t}^{p} = \breve{\mu}_{t+1} - \breve{\sigma}_{t+1}\sqrt{\frac{v-2}{v}}\,\frac{v+\left(t_v^{-1}(p)\right)^2}{v-1}\,\frac{f_{t_v}\left(t_v^{-1}(p)\right)}{p} \tag{11}$$

where f_{t_v} is the probability density function of the Student's t density function with v degree of freedom, defined as follows:

$$f_{t_v}(x) = \frac{\Gamma\left((v+1)/2\right)}{\Gamma(v/2)\sqrt{v\neq}}\left(1+\frac{x^2}{v}\right)^{-(v+1)/2} \tag{12}$$

where $\Gamma(.)$ denotes the Gamma function.

3.2.4. Cornish fisher expansion

The Cornish Fisher expansion is a semi-parametric estimation method that incorporates the skewness and kurtosis of the returns distribution into the VaR and ES forecasts, without any assumption on this distribution. Specifically, the VaR for t+1 is obtained through an extension of the normal quantile function by including the sample skewness ζ_1 and the sample excess kurtosis ζ_2:

$$VaR_{t+1|t}^{p} = \breve{\mu}_{t+1} + \breve{\sigma}_{t+1}CF^{-1}(p) \tag{13}$$

where

$$CF^{-1}(p) = \Phi^{-1}(p) + \frac{\zeta_1}{6}\left(\left(\Phi^{-1}(p)\right)^2 - 1\right)$$
$$+\frac{\zeta_2}{24}\left(\left(\Phi^{-1}(p)\right)^3 - 3\Phi^{-1}(p)\right) - \frac{\zeta_1^2}{36}\left(2\left(\Phi^{-1}(p)\right)^3 - 5\Phi^{-1}(p)\right)$$

The ES for t+1 is determined as follows:

$$ES_{t+1|t}^{p} = \breve{\mu}_{t+1} - \breve{\sigma}_{t+1}ES_{CF}(p) \tag{14}$$

where

$$ES_{CF}(p) = \frac{\varphi\left(\Phi^{-1}(p)\right)}{p}$$
$$\left(1 - \frac{\zeta_2}{24} - \frac{\zeta_1^2}{4} + \frac{\zeta_1}{6}\Phi^{-1}(p) + \left(\frac{\zeta_2}{24} - \frac{\zeta_1^2}{18}\right)\left(\Phi^{-1}(p)\right)^2\right) + \frac{\zeta_1}{6}$$

3.2.5. Exponentially weighted moving average

The methods presented above rest on the moving window of historical returns to capture time-varying in the risk estimates. In financial markets characterized by unexpected shocks and persistence of extreme volatility, known as volatility clustering, weighting all past observations equally (for the historical simulation approach) or using sample moments (for the others approaches) may be insufficient to capture the degree of time-varying and ensure a rapid response to current market conditions. This limited capacity for incorporating conditionality leads to a clustering of VaR violations during turbulent market periods and conversely to unnecessarily higher VaR forecasts during calmer periods (Roncoroni et al., 2015).

To remedy these shortcomings, the generalized autoregressive conditional heteroskedasticity (GARCH) type models or the exponentially weighted moving average volatility (EWMA) model can be considered to estimate the time-varying conditional volatility. The idea behind the EWMA model is to vary the volatility over time to assign a higher weighting to the most recent data. Starting with the sample variance, the conditional variance is chronologically adjusted according to the following formula:

$$\breve{\sigma}_{t+1}^2 = \lambda\breve{\sigma}_t^2 + (1-\lambda)r_t^2 \tag{15}$$

with λ is a smoothing parameter, typically fixed to 0.94 for daily returns.

Assuming that the returns follow a Student's t distribution with zero-mean, the VaR and ES forecasts are then computed as:

$$VaR_{t+1|t}^{p} = \breve{\sigma}_{t+1}\sqrt{v^{-1}(v-2)}t_v^{-1}(p) \tag{16}$$

$$ES_{t+1|t}^{p} = -\breve{\sigma}_{t+1}\sqrt{\frac{v-2}{v}}\,\frac{v+\left(t_v^{-1}(p)\right)^2}{v-1}\,\frac{f_{t_v}\left(t_v^{-1}(p)\right)}{p} \tag{17}$$

3.2.6. GARCH with extreme value theory

A semi-parametric approach proposed by McNeil and Frey (2000) draws on the extreme value theory (EVT) to model the tail of the conditional distribution, regardless the whole distribution of returns. This approach is a two-step procedure that fits firstly a GARCH(1,1) model to the previous n returns to generate the conditional volatility $\breve{\sigma}_{t+1}$ and afterwards estimates the generalized Pareto distribution (GPD) parameters from the negative of the standardized residuals of the GARCH(1,1) model. In particular, the peak-over-threshold method is used to fit the GPD to excesses over a specified threshold.[7]

The probability density function for the GPD with shape parameter $\xi>0$, scale parameter $\beta>0$, and threshold parameter $u\geq0$, is

$$GPD\left(x>u;\xi,\beta\right)=\frac{1}{\beta}\left(1+\xi\frac{x-u}{\beta}\right)^{-1-\frac{1}{\xi}} \tag{18}$$

The quantile can be estimated as:

$$\breve{z}_p=u+\frac{\breve{\beta}}{\breve{\xi}}\left(\left(\frac{n_u}{np}\right)^{\breve{\xi}}-1\right) \tag{19}$$

where n_u is the number of peaks over the threshold.

The VaR and ES forecasts are then obtained as:

$$VaR_{t+1|t}^{p}=-\breve{\sigma}_{t+1}\breve{z}_p \tag{20}$$

$$ES_{t+1|t}^{p}=-\breve{\sigma}_{t+1}\breve{z}_p\left(\frac{1}{1-\breve{\xi}}+\frac{\breve{\beta}-\breve{\xi}u}{\left(1-\breve{\xi}\right)\breve{z}_p}\right) \tag{21}$$

3.3. Evaluating Risk Forecasts Accuracy

The accuracy of risk estimates is typically examined via backtesting which is a technique of model validation based on the comparison of the *ex-ante* risk forecasts from a specific model with the *ex-post* realizations of returns. Below, we briefly describe the most common VaR and ES tests as well as two statistical loss functions used for better comparing the forecasting performance of the different risk models.

3.3.1. VaR tests

To evaluate the performance of the above VaR models, we first considered the unconditional coverage test suggested by Kupiec (1995). This test assesses the unconditional coverage propriety of VaR estimates under which the overall average of VaR violations is not significantly different from its probability level *p*. It compares the observed number of violations to the expected number of violations by testing the hypothesis that the observed violations follow a binomial distribution with parameter *p*. One important weakness of Kupiec's test is that it considers only the frequency of VaR violations and not their magnitude, which may result in an underestimation of risk and lead to very large losses in the worse case scenarios. The risk map proposed by Colletaz et al. (2013) offers a remedy by introducing the concept of super violation, defined as the VaR violations at a very low probability level much smaller than *p*. A likelihood ratio test is applied to assess whether the empirical frequencies of VaR violations and super violations significantly deviate from the theoretical ones. As the two preceding tests do not examine the independence of violations over time, which may cause an acceptation of a model that produces clustered violations, we also consider the conditional coverage test of Christoffersen (1998). Assuming that VaR violations are modelled with a first-order Markov chain, this test examines whether the unconditional coverage and independence properties of correct specified VaR model are jointly fulfilled. In addition to the above suggested tests, we also examine the clustering of violations using the duration test of Christoffersen and Pelletier (2004). According to these authors, this clustering is related to the existence of no-violations durations which are either relatively short by reason of high market volatility or relatively long when the market is calmed down. The authors test the null hypothesis of independence of violations by testing whether the durations are from an exponential distribution, a special case of Weibull distribution under the alternative hypothesis.

3.3.2. ES tests

In order to assess the validity of the ES forecasts, we first consider the exceedance residual test of McNeil and Frey (2000) which is one of the early proposed ES backtests. This test is based on the concept of exceedance residuals, defined as the differences between the actual returns and the ES forecasts conditional on VaR being violated, which should behave like an *i.i.d.* sample from a random variable with mean zero under the null hypothesis of a correctly specified risk model. The authors thus propose to test whether the exceedance residuals have an expected value of zero, where the null distribution of the exceedance residuals is obtained via non-parametric bootstrap. In addition, we consider the conditional calibration test of Nolde and Ziegel (2017), which is based on a Wald-type test statistic that checks whether the expected value of a strict identification function for the VaR and ES is zero. As the two previous tests are formally joint backtests for VaR and ES, we further consider the strict ES regression test of Bayer and Dimitriadis (2020) which only requires ES forecasts as input parameters. That test consists to run a regression of realized returns on the ES forecasts and an intercept term using a linear regression equation. A Wald-type test statistic is used to check whether the intercept and slope parameters equal zero and one, respectively, under the null hypothesis of a correctly specified ES model.[8]

3.3.3. Loss functions

The VaR and ES tests discussed above are mainly meant to evaluate a model adequacy in isolation and not to compare competing models (Acerbi and Szekely, 2014). To remedy this, we consider two loss functions, also known as scoring functions, to compare the forecasting performance of the different risk models: the smaller the average loss given by a scoring function, the better the forecasting model.

A commonly used scoring function for comparing the forecasting accuracy of competing VaR models is the quantile loss function proposed by González-Rivera et al. (2004), also known as the piecewise linear or tick loss function:

$$L_Q=(p-I(r\leq VaR))(r-VaR) \tag{22}$$

Contrary to the VaR, the ES is not an elicitable risk measure and

7 We set the threshold at the 90% quantile of the negative of standardized residuals.

8 For more details on the VaR tests in use, see Hamidi et al. (2015) and Zhang and Nadarajah (2017). For ES tests, refer to Bayer and Dimitriadis (2020) and Hallin and Trucíos (2021).

thus does not admit a strictly consistent scoring function. That problem was overcome by Fissler and Ziegel (2016) by showing that ES is jointly elicitable with VaR. They have thus proposed the following loss function for a joint evaluation of VaR and ES forecasts:

$$L_{FZ} = -\frac{1}{pES} I\left(r \leq VaR\right)\left(VaR - r\right) + \frac{VaR}{ES} + \log\left(-ES\right) - 1 \qquad (23)$$

4. DATA DESCRIPTION

In this study, we consider an index fund that mimics the composition and the performance of the Spanish IBEX 35 index as risky asset. This index is comprised of the 35 most liquid Spanish stocks traded in the continuous market the last 6 months. The high liquidity ensures the execution of portfolio rebalancing operations even during times of increasing volatility that coincide with downturns in market liquidity (Ang et al., 2006b). The futures-based ES targeting strategy is thus implemented using prices of futures contracts on the IBEX 35 index.

The prices are daily closing prices sourced from the official market of these futures, Mercado Español de Futuros Financieros (MEFF), from January 3, 2000 to July 31, 2014. This period was marked by the occurrence of several events, particularly the September attacks in 2001, the bursting of the technology bubble in 2001–2002, the global financial crisis of 2007-2008 and the European sovereign debt crisis of 2010–2012.

In the MEFF, there are at least six futures contracts traded on each trading day.[9] We consider the nearest maturing contracts as they provide the highest liquidity. In regard to rolling a futures position as maturity approaches, we switch to the next-to-nearest maturity contract 10 calendar days before a contract expires to avoid thin markets and expiration effects. Given these implementation choices, the used futures contracts have maturities ranging from 10 through 45 calendar days.

The risk-free rate is the secondary market 1–3 month Spain Treasury bill daily rates taken from the Spanish national central bank website for the same period as that of the futures contracts. Further, a constant dividend yield on the IBEX 35 index is used for each calendar year from 2000 to 2014, for which data was retrieved from Bloomberg. To make one-step ahead VaR and ES forecasts, we follow Taylor (2008), Lazar and Zhang (2019) and Happersberger et al. (2020) and use a rolling window of 1000 past returns to re-estimate parameters for the various risk models on a daily basis. Hence, we consider 4691 daily observations of the IBEX 35 index from December 28, 1995 till July 31, 2014. The index closing values were collected from finance.yahoo.com.

This table reports summary statistics for the IBEX 35 log returns over the period from December 28, 1995 to July 31, 2014. Mean and standard deviation are annualized using a 250-day year. The results of the D'Agostino normality tests for skewness, kurtosis and omnibus are shown in the bottom panel.

9 The number of futures contracts has increased from 6 to 10 on July 2004 and then to 18 on February 2005.

Table 1: Summary and tests statistics of IBEX 35 stock index returns

Mean	**Standard deviation**	**Skewness**	**kurtosis**	**Minimum**	**Maximum**
0.0581	0.2389	0.0039	7.7361	–0.0959	0.1348
D'Agostino-Pearson Normality test		**Skewness**	**Kurtosis**	**Omnibus**	
Statistic		0.1094	22.3511	499.5839	
P-value		0.9129	0.0000	0.0000	

A summary and test statistics for the daily log returns of the IBEX 35 index over the whole time period is provided in Table 1. The annualized return fluctuated around an average of 5.81% and featured a standard deviation of 23.89%. The largest decline in the index was 9.59% in October 10, 2008 at the height of the global financial crisis while the best performance was 13.48% and was recorded on May 10, 2010. Moreover, the D'Agostino-Pearson omnibus test strongly rejects the null hypothesis of normality. In particular, the unconditional return distribution is symmetric around its mean but and exhibits significant excess kurtosis, i.e. extremes tend to be more pronounced than for a normally distributed random variable, reflecting a need for downside protection among investors.

5. EMPIRICAL RESULTS

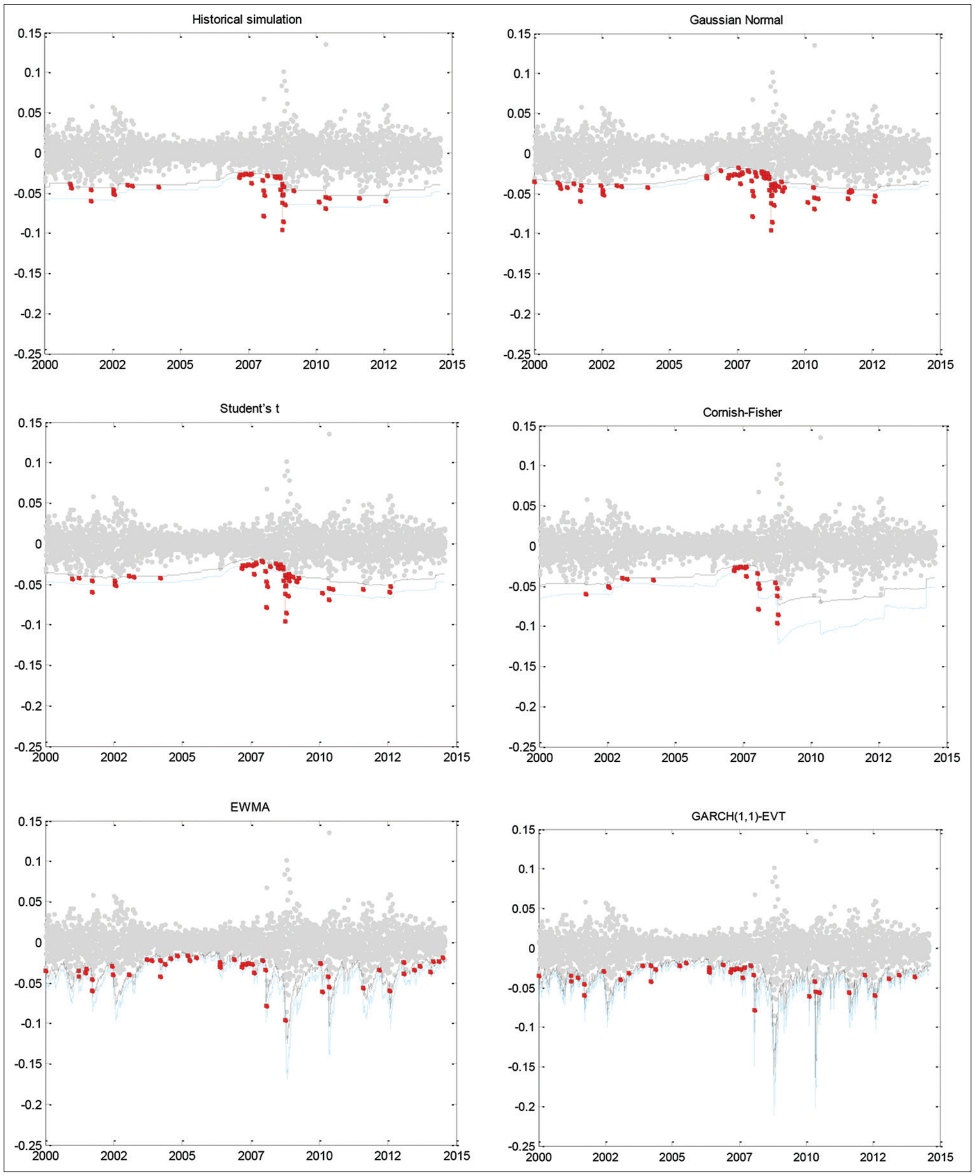
In the following, we evaluate the accuracy of the VaR and ES forecasts made by the different risk models in order to find out the best. Then, we provide evaluations and comparisons of the two procedures of implementing the ES targeting strategy.

5.1. VaR and ES Forecasting Results

Given the time period covered by the futures data, the out-of-sample period consists of 3690 trading days, ranging from January 4, 2000 until July 31, 2014. VaR and ES are forecasted by the risk models discussed above for one trading day ahead in this period using a rolling window of 1000 daily returns at the probability level 1%.

The development of the daily 1% VaR and ES forecasts associated with the different risk models as well as the IBEX 35 returns are depicted in Figure 1. A first glance at the time series of the index returns reveals the typical features of financial assets such as time-varying volatility and volatility clustering. The latter appears notably during times of significant financial market stress. Furthermore, we observe that only the forecasts made the dynamic conditional risk models, EWMA and GARCH(1,1)-EVT, exhibit pronounced variability in time which reflect their substantial response to continuing changes in market conditions. Quite the contrary, unconditional models produce time series of VaR and ES estimates that are extremely smooth (Gaussian Normal and Student's *t* distributions) or flat and characterized by sudden changes on different scales (historical simulation and Cornish-Fisher expansion). This confirms the

Figure 1: VaR and ES forecasts over time



need for conditional variance models in constructing downside risk forecasts. Using only a rolling window of the most recent observations is not enough on its own to deliver acceptable forecasts that capture the degree of time-varying and ensure a rapid response to current market conditions. The consequence is a clustering of VaR violations in time, especially during the global financial market crisis.[10]

The first result from Table 2, which reports the descriptive statistics of the risk estimates, is the broad similarity between the two conditional models EWMA and GARCH(1,1)-EVT. In particular, the VaR and ES forecasts issued from these two models have almost the same mean, standard deviation and maximum values. The only difference is that the GARCH(1,1)-EVT model produced significantly lower forecasts in times of extreme market stress, especially in January and October 2008 and in May 2010. Second, the descriptive statistics confirm the substantial differences between unconditional and conditional risk models. By the use of conditional variance models, the EWMA and GARCH(1,1)-EVT methods generate both the highest, the lowest and the most volatile VaR and ES forecasts. We note furthermore that, for all risk models, ES exhibited greater variability than VaR. This stylized fact confirms the higher sensitivity of ES to changes in market conditions and justify the preference for the ES as a downside risk measure.

This Figure displays the development over the out-of-sample period from January 4, 2000 to July 31, 2014 of the IBEX 35 realized returns (gray dots) as well as the 1%VaR (dashed black line) and the 1%ES (blue line) forecasts made by various risk models using a rolling window of 1000 observations. VaR violations are marked with red squares.

This table reports descriptive statistics for the daily 1% VaR and ES estimates made by various forecasting models over the out-of-

10 A VaR violation is the situation where an *ex-post* negative return is larger, in absolute value, than the VaR.

sample period from January 4, 2000 to July 31, 2014. We report the mean, standard deviation, minimum and maximum

This table reports the results of VaR and ES tests for evaluating 1% VaR and 1% ES forecasts issued from various forecasting models over the out-of-sample period from January 4, 2000 to July 31, 2014. We evaluate the VaR forecasts using the unconditional coverage (UC), the risk map (RM), the conditional coverage (CC) and the duration (DU) test. For testing ES, we consider the exceedance residual (ER), the conditional calibration (CAL) and the strict ES regression (ESR) test. For these ES tests, P-values are given for two-sided hypotheses. We report p-values in bold if >0.10, indicating no evidence against optimality at the 10% significance level. Values between 0.05 and 0.10 are in italics. Furthermore, the second column contains the number of realized VaR violations knowing that expected number of violations is 37 over the whole out-of-sample period. The last two columns report the average loss (scaled by 100) using the quantile loss function (L_Q) and the FZ loss function (L_{FZ}), respectively.

Table 3 gives the number of VaR violations for the various risk models as well as the results for the VaR and ES tests and the average out-of-sample losses based on the FZ loss function and the quantile loss function. Since the expected number of violations is 37, the Cornish-Fisher expansion method clearly overestimates the risk. All the other models produce a number of violations greater than the theoretical value, with an underestimation of risk that is particularly flagrant for the unconditional parametric models, i.e. the Gaussian Normal and Student's *t* models. The GARCH(1,1)-EVT model has the observed number of violations closest the expected one and therefore performs the best in terms of unconditional coverage. The historical simulation approach ranks second, followed by the EWMA model.

According to the P-values from the various VaR tests, the Gaussian Normal, Student's *t* and Cornish-Fisher expansion models fail all the tests due to the large deviation from the expected number of violations. Although the historical simulation approach delivers adequate frequency and magnitude of violations and therefore passes the unconditional coverage and risk map tests, it fails the conditional coverage and duration tests, which shows that these violations are clustered over time. The EWMA model passes the unconditional coverage and risk map tests only at the significance level of 5% because of the large number of violations, and the remainder of VaR test at the significance level of 10%. This gives evidence that the VaR violations occur independently and do not cluster. As expected, the GARCH(1,1)-EVT model performs the best since it is the only model passing all VaR tests at the significance level of 10%.

As can be seen from the ES tests results, the GARCH(1,1)-EVT is once again the only model passing all ES tests at the 10% significance level. The EWMA model fails the conditional calibration test of Nolde and Ziegel (2017). For the other risk models, the ES tests are not meaningful since they fail to provide accurate VaR estimates.

To enable a better comparison of the risk models according to their predictive accuracy of VaR and ES, we further provide in Table 3 the average out-of-sample losses, based on the quantile loss function and the FZ loss function. Clearly, the conditional models, i.e. the EWMA and GARCH(1,1)-EVT, are the most accurate models since they record the lowest values for both loss functions. Particularly, the GARCH(1,1)-EVT is the best-performing model, yielding the lowest average loss based on the FZ loss function. Given the similarity between the average losses from the quantile function, the GARCH(1,1)-EVT model outperforms the EWMA model in predicting ES. Not surprisingly, the unconditional parametric models, Gaussian Normal and Student's *t*, achieve the worst results. This shows that conditionality plays a crucial role in giving accurate VaR and ES estimates.

By these evidences, we conclude that the GARCH(1,1)-EVT model delivers the most accurate VaR and ES estimates, enabling reliable capture of the stylized facts observed for the IBEX 35 index returns such as time-varying volatility, non-normality, fat

Table 2: Descriptive statistics for out-of-sample VaR and ES forecasts

	Mean		Standard deviation		Minimum		Maximum	
	VaR	ES	VaR	ES	VaR	ES	VaR	ES
Historical simulation	–0.041	–0.053	0.008	0.011	–0.053	–0.068	–0.022	–0.029
Gaussian Normal	–0.035	–0.040	0.007	0.008	–0.047	–0.054	–0.018	–0.020
Student's *t*	–0.039	–0.050	0.008	0.011	–0.052	–0.068	–0.019	–0.025
Cornish-Fisher	–0.048	–0.067	0.013	0.024	–0.074	–0.123	–0.023	–0.029
EWMA	–0.036	–0.046	0.017	0.023	–0.126	–0.170	–0.012	–0.015
GARCH (1,1)-EVT	–0.037	–0.045	0.018	0.022	–0.164	–0.212	–0.013	–0.017

Table 3: Backtesting on out-of-sample VaR and ES forecasts

	VaR tests					ES tests			Average loss	
	Viol	UC	RM	DU	CC	ER	CAL	ESR	L_Q	L_{FZ}
Historical simulation	46	**0.15**	**0.13**	0.00	0.00	**0.51**	**0.40**	*0.05*	0.055	1.738
Gaussian Normal	82	0.00	0.00	0.00	0.00	0.00	0.00	*0.05*	0.060	2.054
Student's *t*	61	0.00	0.00	0.00	0.00	**0.43**	0.01	0.01	0.057	1.846
Cornish-Fisher	22	0.01	0.03	0.00	0.03	**0.58**	0.00	0.00	0.054	1.719
EWMA	49	*0.06*	*0.08*	**0.17**	**0.15**	**0.93**	0.04	**0.25**	0.045	1.502
GARCH (1,1)-EVT	39	**0.73**	**0.91**	**0.10**	**0.68**	**0.83**	**0.44**	**0.63**	0.044	1.446

tails and volatility clustering, and therefore a quick reaction to the prevailing risk environment. This latter feature is crucial for the tail risk protection strategies since wrong risk timing translates into a high exposure to the risky asset when the market's downside risk is high and vice versa.

5.2. Outcome of the ES Targeting Strategy

In the light of the above results, the ES forecasts at 1% made by the GARCH(1,1)-EVT model serve as basis for the risk targeting strategy implementation over the out-of-sample period. To ensure a perfect downside protection, we first set the target ES at a low level equal to 1.5% and rebalance the portfolios on a daily basis. We also integrate the following restrictions and assumptions to create a realistic investment environment. For the index-and-bill implementation, we impose constraints on leverage and short selling and assume transaction costs of 10 basis points applied on the portfolio adjustment size. For the futures-based implementation, we impose constraint on long futures position at the end of every trading day and assume transaction costs of 2 basis points applied on the position adjustment size. We further exclude margin requirements and consider that daily mark-to-market gains/losses are neither invested nor financed. Finally, the transaction costs are charged on the transaction date and settled on the first trading day thereafter.

The performances of the two procedures of implementing the ES targeting strategy are compared on the basis of descriptive statistics, protection effectiveness and risk-adjusted performance. The descriptive statistics are the mean, standard deviation, skewness and kurtosis. The protection effectiveness measures are the maximum drawdown and ES with significance level at 1%.[11] The risk-adjusted performance is evaluated with the help of the Sortino, Omega, Calmar and upside potential ratios, that have been considered most relevant for the target risk strategies (Annaert et al., 2009, Bertrand and Prigent, 2011). Sortino ratio measures excess return over a predefined target return per unit of downside deviation.[12] Omega ratio is computed by dividing the higher partial moment of degree one by the lower partial moment of the same degree, that is, the expected excess return over a predefined threshold divided by the expected loss below the same threshold. Calmar ratio is the annualized mean return relative to the maximum drawdown. The upside potential ratio is the higher partial moment of degree one divided by the downside deviation. Sortino and Calmar ratios are refinements to the popular Sharpe ratio that better addresses the risk preferences of investors (risk-seeking above a certain threshold return and risk-averse below this threshold). They favour investments with the highest return but the least downside deviation for the former and the least maximum drawdown for the latter. The upside potential ratio is an alternative to the Sortino that has the advantage of using the minimum acceptable return for evaluating both profits and losses (Plantinga and de Groot, 2001). It favours investments which have had relatively better upside performance per unit of downside risk.

11 The maximum drawdown represents the maximum percentage loss that a portfolio incurred from its peak level to its lowest level over a given period of time.

12 The downside deviation measures the standard deviation of the only returns below a minimum target return (here, zero).

The main benefit of the Calmar ratio is the consideration of the entire return distribution without imposing any parametric shape, which allows the capture of information in the higher moments.

This chart illustrates the performance over the out-of-sample period of the risk targeting strategy with daily rebalancing under two implementation procedures: index-and-bill (black line) and index and futures (red line). The target level is a 1.5% ES. For comparison, we include the performance of the buy-and hold index investment strategy (blue line). The related transaction costs are included in the valuation. Panel (a) shows the cumulative returns of investing 1 Euro in January 3, 2000 until July 31, 2014. Panel (b) shows the kernel densities of the daily returns for the strategies.

For a visual representation of the difference in performance between the index-and-bill implementation and the index-and-futures implementation, Figure 2a plots the cumulative returns of investing 1 Euro in each portfolio over the out-of-sample period. For comparison, the performance of the buy-and-hold index investment strategy over the same period is also shown. As can be seen, the evolution is globally similar under the two implementation procedures. An apparent discrepancy emerged nevertheless from the beginning of 2007 when the cumulative value of the futures-based strategy becomes clearly higher. This observation must not however lead to the conclusion that the futures-based ES targeting strategy dominates its index-and-bill rival because the outperformance may depend on the ES target level and/or the rebalancing frequency.

Irrespective of the implementation procedure, the ES targeting strategy succeeded in preserving the invested capital, capturing the upside potential of the underlying and protecting the downside especially during the crises periods. With the exception of the year 2007 and the second quarter of the year 2014, during which the investment in the market index has reached record levels, the values of the protected portfolios exceed that of the buy-and-hold portfolio. This performance is mainly driven by the ability of the risk targeting strategy to mitigate the repercussions of the various crises occurring during the considered period.

Figure 2b plots the kernel densities of the simulated returns of the protected portfolios in comparison with the unprotected investment strategy in the equity index. Clearly, the ES targeting strategy has provided major shift of mass in the distribution from the tails towards the center, resulting in significant reduction in tail risk. As with all protection strategies, ES targeting incurs an implicit cost in that highly positive returns can no longer achievable is case of upward movements of the underlying index. While the daily returns vary between –9.12% and 14.48% for the underlying index, they only lie within the range of –2.59% and 2.24% for the index-and-bill implementation and the range of –2.36% and 2.15% for the futures-based implementation.

This table summarizes the estimation results after and before transaction costs over the out-of-sample period for the two implementation procedures of the ES targeting strategy: index-and bill and futures-based. For comparison, we include the performance of the underlying index as well as the estimation results for the futures-based implementation based on theoretical futures prices. We report the annualized mean return (Mean), annualized standard deviation (SD), skewness (Skew), kurtosis (Kurt), maximum drawdown (MDD), 1% ES, Sortino ratio (SoR), Omega ratio, Calmar ratio and upside potential ratio (UPR). To calculate the Sortino and Omega ratios, the MAR is set to zero. Apart from skewness, kurtosis and Omega ratio, all statistics are expressed as percentages. The statistics are presented in panel A when transaction costs are included, and in panel B no transaction costs are accounted for.

Table 4 reports the corresponding backtesting results of the ES targeting strategy under its two implementation procedures as well as the performance of the underlying index. Panel A shows the results when transaction costs are included. First of all, we note that ES of the protected portfolios are very closely to the pre-specified ES target level of 1.5%, reflecting the high precision of both implementation procedures. Further, it appears that the drag on performance induced by the ES targeting strategy not only caused a major drop in mean return but also negatively affected the asymmetry of the return distribution. Indeed, the skewness has changed sign to negative given that the range of forgone positive returns is much broader than the discarded negative returns. This result contradicts findings by Dreyer and Hubrich (2019) that managed volatility strategies deliver an enhanced skewness. In exchange of this protection cost, the strategy contributes to substantial reductions in the standard deviation and kurtosis of the return distribution. Moreover, the significant falls in maximum drawdown reflect the strategy effectiveness in mitigating extremely negative returns. Accordingly, the ES targeting strategy leads to huge improvements in all risk-adjusted return ratios compared to the underlying owing to the substantial downside risk reduction.

Comparing across the two implementation procedures, the results reported in Panel A of Table 4 clearly indicate that implementing ES targeting strategy using futures contracts leads to better drawdown protection and risk-adjusted performance. In fact, the futures-based implementation delivers a lower maximum drawdown that its index-and-bill counterpart (19.90% compared to 21.66%), and generates returns that are less negatively skewed and less leptokurtic in distribution. It also produces higher annualized mean return (3.49% compared to 2.95%) and thus better performance based on all risk-adjusted return ratios.

To check whether this evidence is attributable to the transaction cost advantage, we have made the same comparison between the two implementation procedures assuming that no transaction costs are charged. Panel B of Table 4 shows the corresponding results. We see that the standard deviation and higher moments of the return distribution as well as the 1% ES are the same as when taking into account transaction costs. In addition, significant improvement in mean return (from 2.95% to 3.83%) and maximum drawdown (from 21.66% to 18.83%) can be observed for the index-and-bill implementation compared to slightly improvements for the futures-based implementation given the relatively small transaction costs. The index-and-bill implementation takes thus the lead over its futures-based rival in terms of protection effectiveness and risk-adjusted return ratios, except for the upside potential ratio. Two important conclusions can be drawn from the results. First, the transaction costs tend to shift the return distribution of the ES targeting strategy to the left without changing its shape, which translates into lower mean return and higher maximum drawdown. This particularly affects the performance and the protection effectiveness of the index-and-bill implementation because of the importance of transaction costs. Second, the fact

Figure 2: (a and b) Performance of the two procedures of implementing the ES targeting strategy

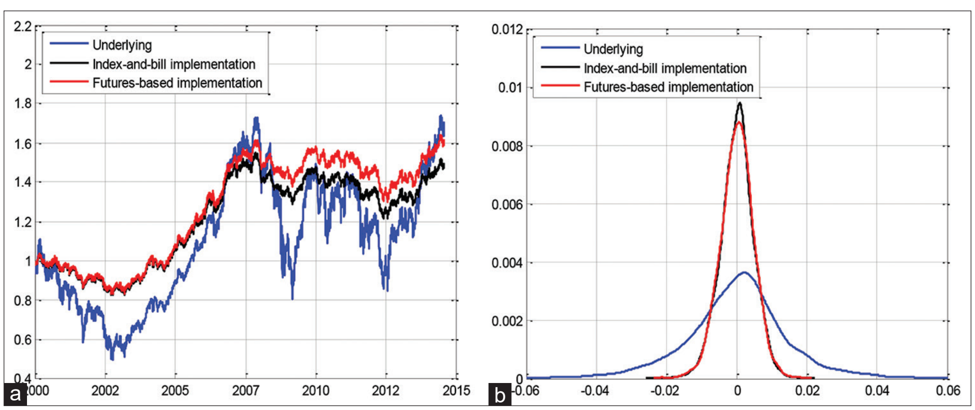


Table 4: Backtest performance of risk targeting strategy

	Mean	SD	Skew	Kurt	MDD	ES	SoR	Omega	Calmar	UPR
Underlying	6.42	24.22	0.27	8.49	55.46	-5.19	2.41	1.048	11.57	52.3
Panel A: Including transaction costs										
Index-and-bill	2.95	7.50	-0.21	4.01	21.66	-1.50	3.50	1.067	13.63	56.1
Futures based	3.49	7.58	-0.17	3.87	19.90	-1.50	4.12	1.078	17.53	57.2
Futures based using theoretical prices	3.30	7.53	-0.21	4.00	20.94	-1.51	3.91	1.074	15.75	56.5
Panel B: No transaction costs										
Index-and-bill	3.83	7.50	-0.21	4.01	18.83	-1.50	4.56	1.087	20.32	57.0
Futures based	3.57	7.57	-0.17	3.87	19.75	-1.50	4.22	1.080	18.06	57.2
Futures based using theoretical prices	3.38	7.53	-0.21	4.00	20.63	-1.51	4.01	1.076	16.39	56.6

that the futures-based implementation generates the highest upside potential ratio even when no transaction costs are accounted for suggests that the transaction cost advantage cannot be the only source of its outperformance. Indeed, it seems that the replication imperfections due to futures mispricing gives the futures-based implementation a better participation in upward movements of the underlying index.

To check this, we follow Merrick (1988) and Do and Faff (2004) and perform additional simulations for the futures-based implementation, with and without transaction costs, using the theoretical futures prices as determined by the cost-of-carry model instead of the actual futures prices. The last row of both Panels A and B in Table 4 provides the corresponding results. We observe that the simulation based on theoretical futures prices leads practically to the same standard deviation, skewness, kurtosis and ES as the index-and-bill implementation, confirming that the theoretical futures prices derived from the cost-of-carry model allow replicating the statistical proprieties of the ES targeting strategy. Unreported chart shows that the corresponding kernel densities of the simulated daily returns are coincident. Notably, the simulation using theoretical futures prices performs significantly worse than the simulation based on actual futures prices both in terms of effectiveness of protection and risk-adjusted performance, with transaction costs either included or not included. This finding can be explained by the fact that the index futures mispricing can either enhance or deteriorate the short-term performance of the strategy (Merrick, 1988) but has a long-term beneficial effect. So that, our empirical results do not contradict the empirical findings of Do and Faff (2004) as they examined the portfolio insurance strategies for subsamples of 3-month maturity. As a conclusion, the superior performance of the futures-based implementation relative to its index-and-bill rival is driven both by the transaction cost advantage and by the long-term beneficial effect of the futures mispricing that allows a higher capacity to capture the upside potential of the underlying index.

To put into perspective this difference between the two implementation methods of the ES targeting strategy, additional simulations are performed to examine their sensitivities to the risk target level and rebalancing frequency. In this respect, we regenerate the simulations with daily rebalancing at the following ES target levels: 1%, 2%, and 2.5%. Subsequently, we target an ES level of 1.5% and apply three rebalancing triggers. In particular, we impose that rebalancing trades are only triggered when the absolute dividend-adjusted return of the underlying index is ≥1%, 1.5% and 2%, successively.[13] The corresponding results are summarized in Panel A and Panel B of Table 5, respectively.

Results in Panel A show that the increase in the risk target level makes the strategy vulnerable to leverage constraint for the index-and-bill implementation and short-only constraint for the futures-based implementation. This limits exposure to risk and brings the absolute ES of the portfolios slightly below the targeted level. In particular, the effect of implementation constraints only becomes apparent when an ES level of 2.5% is targeted but remains marginal due to daily rebalancing. As well, although higher moments of the simulated returns exhibit low sensitivity towards risk target level, we can see a narrowing of skewness and kurtosis spreads between the two implementation methods by increasing the target level.

In addition, it appears that the ES targeting strategy generates better risk-adjusted performance at lower ES target levels, which is in line with Rickenberg (2020) and Happersberger et al. (2020). Higher level of target risk significantly increased the turnover of the index-and-bill implementation, which translates into greater transaction costs and reinforces the transaction cost advantage of the futures-based implementation. As a consequence, the latter records the highest mean returns and the best performance measures at all risk levels.

This table summarizes the estimation results over the out-of-sample period for the two implementation methods of the ES targeting strategy using various ES target levels (Panel A) and various trigger-based rebalancing levels(Panel B). We report the annualized mean return (Mean), annualized standard deviation (SD), skewness (Skew), kurtosis (Kurt), maximum drawdown (MDD), 1% ES, Sortino ratio (SoR), Omega ratio, Calmar ratio and upside potential ratio (UPR). Sortino and Omega ratios are computed using a minimum accept return of zero. Apart from skewness, kurtosis and Omega ratio, all statistics are expressed as percentages.

The futures-based implementation also outperforms its index-and-bill counterpart in terms of protection effectiveness by delivering lower absolute ES and maximum drawdown at all ES target levels. Due to the transaction cost effect, the gap between the two implementation procedures is more pronounced at higher ES target levels. On the whole, these findings support the ability of the futures-based implementation to produce superior risk-adjusted performance and a better downside protection at all ES target levels.

Furthermore, results in Panel B reveal that the application of a threshold-based rebalancing reduces not only the transaction costs but also the exposure to risk. As a result, the absolute ES of the portfolios significantly fell below the ES target level of 1.5%. The maximum drawdown is likewise reduced, which means that ES targeting strategy offers better protection effectiveness when less frequent rebalancing intervals are used. This enhancement in downside protection was accompanied by a notable improvement in risk-adjusted performance of the index-and-bill implementation proportionally with the decline in the transaction costs. The performance improvement is however of less importance for the futures-based implementation and approximately the same throughout the different trigger levels.

By comparing the results obtained with the two implementation procedures, it turns out that reducing the rebalancing frequency by

13 The trigger levels are chosen on the basis of observed daily dividend-adjusted returns of the index. Thus, the 2% trigger level entails adjusting the portfolio consisting of index and bill 552 times during the out-of-sample period (14.96%) and 701 times for the futures position throughout the same period (19%). This difference is due to the fact that the position in futures contracts must be readjusted at each rollover date irrespective to the index return. Given the above percentages, a higher trigger lever would be meaningless.

Table 5: Risk targeting: daily rebalancing and various target levels

	Mean	SD	Skew	Kurt	MDD	ES	SoR	Omega	Calmar	UPR
Panel A: Daily rebalancing and various target levels										
1% ES Target										
Index-and-bill	2.70	5.00	–0.20	4.01	12.63	–1.00	4.84	1.092	21.41	57.2
Futures based	3.04	5.23	–0.16	3.83	11.92	–1.04	5.23	1.099	25.48	58.0
1.5% ES Target										
Index-and-bill	2.95	7.50	–0.21	4.01	21.66	–1.50	3.50	1.067	13.63	56.1
Futures based	3.49	7.58	–0.17	3.87	19.90	–1.50	4.12	1.078	17.53	57.2
2% ES Target										
Index-and-bill	3.25	9.99	–0.20	3.99	29.97	–2.00	2.89	1.055	10.85	55.7
Futures based	3.89	9.93	–0.18	3.90	28.01	–1.96	3.49	1.066	13.88	56.6
2.5% ES Target										
Index-and-bill	3.39	12.29	–0.20	3.97	37.57	–2.44	2.44	1.046	9.02	55.3
Futures based	3.92	12.16	–0.18	3.96	35.53	–2.40	2.86	1.054	11.04	55.8
Panel B: Target ES of 1.5% and various rebalancing triggers										
Daily rebalancing										
Index-and-bill	2.95	7.50	–0.21	4.01	21.66	–1.50	3.50	1.067	13.63	56.1
Futures based	3.49	7.58	–0.17	3.87	19.90	–1.50	4.12	1.078	17.53	57.2
1% Trigger level										
Index-and-bill	3.37	7.06	–0.17	3.96	18.00	–1.39	4.27	1.081	18.70	56.9
Futures based	3.52	7.23	–0.15	3.85	18.19	–1.42	4.37	1.082	19.35	57.4
1.5% Trigger level										
Index-and-bill	3.40	6.76	–0.15	4.08	16.81	–1.33	4.52	1.086	20.26	57.0
Futures based	3.49	7.12	–0.15	3.95	17.88	–1.40	4.41	1.083	19.54	57.3
2% Trigger level										
Index-and-bill	3.34	6.43	-0.16	4.18	15.92	–1.27	4.66	1.089	20.96	56.9
Futures based	3.54	7.08	–0.15	4.03	18.06	–1.41	4.50	1.085	19.60	57.3

applying a trigger threshold of 1% is enough to make the index-and-bill implementation more effective at protecting the downside risk than its futures-based counterpart. This effectiveness was associated with improved performance. In fact, the index-and-bill implementation generates the highest Omega, Sortino and Calmar ratios from the trigger level of 1.5%, but still fail to outperform its futures-based rival in terms of mean return and upside potential ratio. This in turn causes a severe drag on performance during upward market trends which is unfavourable from an investor's perspective.

6. CONCLUSION

Recurrent financial crises in recent years have highlighted the increasingly need for tail risk hedging strategies to protect against a repeat of such shocks in the future. This study provides a comprehensive evidence of the benefits of using stock index futures to manage downside risk. We concentrated our analysis on comparing the relative performance of two methods for implementing ES targeting strategy. On one hand, the index-and-bill implementation aims to keep the ES of the portfolio equal to a pre-specified target level by dynamically shifting money between the risky and the riskless asset based on updated forecasts of the risky asset's ES. On the other hand, the futures-based implementation attempts to achieve the same end by simply short futures written on the risky asset and continually adjust the futures position according to the ES forecasts.

Using 1% ES forecasts issued from the GARCH(1,1)-EVT model, which is the best-performing model among a set of competing approaches according to the most common VaR and ES tests and loss functions, our empirical analysis confirms that the ES targeting strategy successfully capture the upside potential of the market and reduce substantially the downside risk, leading to an enhanced risk-adjusted performance compared a buy-and-hold index investment strategy. With respect to the return distribution, the ES targeting strategy tends to incur implicit cost in the form of decreases in mean and skewness in exchange for reduced standard deviation and kurtosis. Furthermore, it leads to better results when a low ES target level is chosen and a high trigger level is applied for rebalancing.

When comparing results across the two implementation procedures, it turns out that the futures-based implementation in general outperforms its index-and-bill rival in terms of both risk-adjusted performance and effectiveness of downside protection. This outperformance is driven not only by the transaction cost advantage in the futures markets, but also by the replication imperfections due to futures mispricing which offer in the long term better participation in upward movements of the underlying index.

As a robustness check, additional simulations for the ES targeting strategy are performed using various ES target levels and rebalancing triggers. While the outperformance of the futures-based implementation over its index-and-bill counterpart is confirmed for all target levels, the application of rebalancing triggers offers an enhanced risk-adjusted performance and a better downside protection but does not have the same effect on the two implementation procedures: the futures-based implementation remains insensitive to changes in the trigger level whereas the index-and-bill implementation succeeds to generate the best results both in terms of protection effectiveness and higher Sortino, Omega and Calmar ratios when high trigger level is used. However, while the index-and-bill implementation benefits substantially from lower transaction costs as a result of this threshold based rebalancing, the significant drag on performance relative to the futures-based implementation affects its attractiveness for investors.

REFERENCES

Acerbi, C., Szekely, B. (2014), Backtesting expected shortfall. Risk, 27(11), 76-81.

Andersen, T.G., Bollerslev, T., Christoffersen, P., Diebold, F.X. (2011), Financial risk measurement for financial risk management. SSRN Electronic Journal, 2, 1127-1220.

Ang, A., Chen, J., Xing, Y. (2006a), Downside risk. Review of Financial Studies, 19(4), 1191-1239.

Ang, A., Hodrick, R.J., Xing, Y., Zhang, X. (2006b), The cross-section of volatility and expected returns. The Journal of Finance, 61(1), 259-299.

Annaert, J., van Osselaer, S., Verstraete, B. (2007), Performance evaluation of portfolio insurance strategies using stochastic dominance criteria. SSRN Electronic Journal, 33(2), 272-280.

Artzner, P., Delbaen, F., Eber, J.M., Heath, D. (1999), Coherent measures of risk. Mathematical Finance, 9(3), 203-228.

Bali, T.G., Demirtas, K.O., Levy, H. (2009), Is there an intertemporal relation between downside risk and expected returns? Journal of Financial and Quantitative Analysis, 44(4), 883-909.

Basak, S., Shapiro, A. (2001), Value-at-risk-based risk management: Optimal policies and asset prices. Review of Financial Studies, 14(2), 371-405.

Bayer, S., Dimitriadis, T. (2020), Regression-based expected shortfall backtesting. Journal of Financial Econometrics, (forthcoming), 2020, nbaa013.

Benson, R., Shapiro, R.K., Smith, D., Thomas, R. (2013), A comparison of tail risk protection strategies in the U.S. market. Alternative Investment Analyst Review, 17, 32-47.

Bertrand, P., Prigent, J.L. (2011), Omega performance measure and portfolio insurance. Journal of Banking and Finance, 35(7), 1811-1823.

Bhansali, V. (2011), Beyond risk parity. The Journal of Investing, 20(1), 137-147.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L.H. (2018), Risk everywhere: Modeling and managing volatility. The Review of Financial Studies, 31(7), 2729-2773.

Bollerslev, T., Todorov, V., Xu, L. (2015), Tail risk premia and return predictability. Journal of Financial Economics, 118(1), 113-134.

Bongaerts, D., Kang, X., van Dijk, M. (2020), Conditional volatility targeting. Financial Analysts Journal, 76(4), 54-71.

Buraschi, A., Porchia, P., Trojani, F. (2010), Correlation risk and optimal portfolio choice. The Journal of Finance, 65(1), 393-420.

Campbell, R., Koedijk, K., Kofman, P. (2002), Increased correlation in bear markets. Financial Analysts Journal, 58(1), 87-94.

Chabi-Yo, F., Ruenzi, S., Weigert, F. (2018), Crash sensitivity and the cross section of expected stock returns. Journal of Financial and Quantitative Analysis, 53(3), 1059-1100.

Christoffersen, P. (2016), Elements of Financial Risk Management. 2nd ed. United States: Academic Press.

Christoffersen, P., Errunza, V., Jacobs, K., Langlois, H. (2012), Is the potential for international diversification disappearing? A dynamic copula approach. Review of Financial Studies, 25(12), 3711-3751.

Christoffersen, P., Pelletier, D. (2004), Backtesting value-at-risk: A duration-based approach. Journal of Financial Econometrics, 2(1), 84-108.

Christoffersen, P.F. (1998), Evaluating interval forecasts. International Economic Review, 39(4), 841.

Colletaz, G., Hurlin, C., Pérignon, C. (2013), The risk map: A new tool for validating risk models. Journal of Banking and Finance, 37(10), 3843-3854.

Cooper, T. (2010), Alpha generation and risk smoothing using managed volatility. SSRN Electronic Journal, 2010, 1664823.

Dachraoui, K. (2018), On the optimality of target volatility strategies. The Journal of Portfolio Management, 44(5), 58-67.

di Persio, L., Garbelli, M., Wallbaum, K. (2021), Forward-looking volatility estimation for risk-managed investment strategies during the COVID-19 crisis. Risks, 9(2), 33-40.

Do, B.H., Faff, R.W. (2004), Do futures-based strategies enhance dynamic portfolio insurance? Journal of Futures Markets, 24(6), 591-608.

Dowd, K. (2005), Measuring Market Risk. 2nd ed. United States: Wiley.

Dreyer, A.A., Hubrich, S. (2019), Tail-risk mitigation with managed volatility strategies. Journal of Investment Strategies, 8(1), 29-56.

Engle, R.F., Manganelli, S. (2004), CAViaR: Conditional autoregressive value at risk by regression quantiles. Journal of Business and Economic Statistics, 22(4), 367-381.

Fissler, T., Ziegel, J.F. (2016), Higher order elicitability and Osband's principle. The Annals of Statistics, 44(4), 1680-1707.

Giese, G. (2012), Optimal design of volatility-driven algo-alpha trading strategies. Risk, 25(6), 68-73.

González-Rivera, G., Lee, T.H., Mishra, S. (2004), Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. International Journal of Forecasting, 20(4), 629-645.

Gormsen, N.J., Jensen, C.S. (2020), Higher-moment Risk, Working Paper. Illinois: University of Chicago Booth Schools of Business and Bocconi University.

Hallerbach, W.G. (2012), A proof of the optimality of volatility weighting over time. The Journal of Investment Strategies, 1(4), 87-99.

Hallerbach, W.G. (2015), Advances in portfolio risk control. In: Jurczenko, E., editor. Risk-based and Factor Investing. Amsterdam: Elsevier. p1-30.

Hallin, M., Trucíos, C. (2021), Forecasting value-at-risk and expected shortfall in large portfolios: A general dynamic factor model approach. Econometrics and Statistics, (forthcoming), 2021, 55205.

Hamidi, B., Hurlin, C., Kouontchou, P., Maillet, B. (2015), A DARE for VaR. Finance, 36(1), 7-38.

Happersberger, D., Lohre, H., Nolte, I. (2020), Estimating portfolio risk for tail risk protection strategies. European Financial Management, 26(4), 1107-1146.

Harvey, C.R., Hoyle, E., Korgaonkar, R., Rattray, S., Sargaison, M., Hemert, O.V. (2018), The impact of volatility targeting. The Journal of Portfolio Management, 45(1), 14-33.

Hocquard, A., Ng, S., Papageorgiou, N. (2013), A constant-volatility framework for managing tail risk. The Journal of Portfolio Management, 39(2), 28-40.

Ilmanen, A., Kizer, J. (2012), The death of diversification has been greatly exaggerated. Journal of Portfolio Management, 38(3), 15-27.

Jondeau, E., Rockinger, M. (2003), Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. Journal of Economic Dynamics and Control, 27(10), 1699-1737.

Jorion, P. (2021), Financial Risk Manager Handbook. 5th ed. United States: Wiley.

Kirby, C., Ostdiek, B. (2012), It's all in the timing: Simple active portfolio strategies that outperform naïve diversification. Journal of Financial and Quantitative Analysis, 47(2), 437-467.

Kuester, K., Mittnik, S., Paolella, M.S. (2006), Value-at-risk prediction: A comparison of alternative strategies. Journal of Financial Econometrics, 4(1), 53-89.

Kupiec, P.H. (1995), Techniques for verifying the accuracy of risk measurement models. The Journal of Derivatives, 3(2), 73-84.

Lazar, E., Zhang, N. (2019), Model risk of expected shortfall. Journal of Banking and Finance, 105, 74-93.

Longin, F., Solnik, B. (2001), Extreme correlation of international equity markets. The Journal of Finance, 56(2), 649-676.

Loria, S., Pham, T.M., Sim, A.B. (1991), The performance of a stock index futures based portfolio insurance scheme: Australia evidence. Review of Futures Markets, 10, 438-457.

McNeil, A.J., Frey, R. (2000), Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. Journal of Empirical Finance, 7(3), 271-300.

Merrick, J.J. (1988), Portfolio insurance with stock index futures. Journal of Futures Markets, 8(4), 441-455.

Moreira, A., Muir, T. (2017), Volatility-managed portfolios. The Journal of Finance, 72(4), 1611-1644.

Nolde, N., Ziegel, J.F. (2017), Elicitability and backtesting: Perspectives for banking regulation. The Annals of Applied Statistics, 11(4), 1833-1874.

Papageorgiou, N.A., Reeves, J.J., Sherris, M. (2017), Equity Investing with Targeted Constant Volatility exposure (January 1, 2017). FIRN Research Paper No. 2614828, UNSW Business School Research Paper. Available from: https://www.ssrn.com/abstract=2614828

Perchet, R., de Carvalho, R.L., Heckel, T., Moulin, P. (2016), Predicting the success of volatility targeting strategies: Application to equities and other asset classes. The Journal of Alternative Investments, 18(3), 21-38.

Plantinga, A., de Groot, S. (2001), Risk-adjusted performance measures and implied risk-attitudes. SSRN Electronic Journal, 2001, 289193.

Poon, S.H., Rockinger, M., Tawn, J. (2004), Extreme value dependence in financial markets: Diagnostics, models, and financial implications. Review of Financial Studies, 17(2), 581-610.

Rickenberg, L. (2020), Tail risk targeting: Target VaR and CVaR strategies. SSRN Electronic Journal, 2020, 3444999.

Righi, M.B., Ceretta, P.S. (2015), A comparison of expected shortfall estimation models. Journal of Economics and Business, 78, 14-47.

Roncoroni, A., Fusai, G., Cummins, M. (2015), Handbook of Multi-Commodity Markets and Products: Structuring, Trading and Risk Management. 1st ed. United States: Wiley.

Strub, I.S. (2013), Tail Hedging Strategies (May 7, 2016). Available from: https://www.ssrn.com/abstract=2261831

Taylor, J.W. (2008), Estimating value at risk and expected shortfall using expectiles. Journal of Financial Econometrics, 6(2), 231-252.

Wang, P., Sullivan, R.N., Ge, Y. (2012), Risk-based dynamic asset allocation with extreme tails and correlations. The Journal of Portfolio Management, 38(4), 26-42.

Yamai, Y., Yoshiba, T. (2002), On the validity of value-at-risk: Comparative analyses with expected shortfall. Monetary and Economic Studies, 20(1), 57-85.

Zhang, Y., Nadarajah, S. (2017), A review of backtesting for value at risk. Communications in Statistics-Theory and Methods, 47(15), 3616-3639.

APPENDIX

Proof of Equation (2)

We consider the following two portfolios:

Portfolio 1: a zero-coupon bond with a face value of B_t.

Portfolio 2: $\frac{B_t}{S_t}$ shares of the risky asset plus short $n_t = \frac{B_t}{S_t e^{(r_{f,t}-q_t)(T-t)-\frac{r_{f,t}}{365}}}$ futures contracts written on the risky asset with time to maturity $T–t$.

The following day t+1, portfolio 1 is worth $B_t e^{\frac{r_{f,t}}{365}}$. In portfolio 2, the shares will be worth $\frac{B_t}{S_t} S_{t+1} e^{\frac{q_t}{365}}$. The gain/loss from the short futures position is:

$$\frac{B_t}{S_t e^{(r_{f,t}-q_t)(T-t)-\frac{r_{f,t}}{365}}}\left(F_t - F_{t+1}\right) = \frac{B_t}{S_t e^{(r_{f,t}-q_t)(T-t)-\frac{r_{f,t}}{365}}}\left(S_t e^{(r_{f,t}-q_t)(T-t)} - S_{t+1} e^{(r_{f,t+1}-q_{t+1})\left(T-t-\frac{1}{365}\right)}\right)$$

For daily frequency of portfolio rebalancing, we may write as approximations that $r_{f,t+1} \approx r_{f,t}$ and $q_{t+1} \approx q_t$. In this case, the gain/loss reduces to $B_t e^{\frac{r_{f,t}}{365}} - \frac{B_t}{S_t} S_{t+1} e^{\frac{q_t}{365}}$.

In conclusion, the components of portfolio 2 is worth $B_t e^{\frac{r_{f,t}}{365}}$ at time t+1, which is identical to portfolio 1.