Coherent risk measures under filtered historical simulation

Kostas Giannopoulos a,*, Radu Tunaru b

a UAE University, CBE P.O. Box 17555 Al Ain, AD, United Arab Emirates
b London Metropolitan University, 31 Jewry Street, London EC3N 2EY, UK

Available online 29 September 2004

Abstract

Recent studies have strongly criticised conventional VaR models for not providing a coherent risk measure. Acerbi provides the intuition for an entire family of coherent measures of risk known as “spectral risk measures” [Spectral measures of risk: A coherent representation of subjective risk aversion. Journal of Banking and Finance 26 (7) (2002) 1505–1518]. In this study we illustrate how the Filtered Historical Simulation [Barone-Adesi, G., Bourgoin, F., Giannopoulos, K., 1998. Don’t look back. Risk 11, 100–104; Barone-Adesi, Giannopoulos, K., Vosper, L., 1999. VaR without correlations for non-linear portfolios. Journal of Futures Markets 19, 583–602], can provide an improved methodology for calculating the Expected Shortfall. Thereafter, we prove that these new risk measures are spectral and are coherent as well, following Acerbi. Furthermore, we provide the statistical error formula that allows to calculate the error for our model.

© 2004 Elsevier B.V. All rights reserved.

JEL classification: D81; C13
Keywords: Expected shortfall; Filtered historical simulation; Ordered statistics; Coherent measure; Spectral risk measure; Generalised extreme value distribution

* Corresponding author.
E-mail address: kgiannopoulos@uaeu.ac.ae (K. Giannopoulos).
1. Introduction

When Value-at-Risk (VaR) was first introduced it achieved great popularity among regulators and financial institutions. VaR is an estimator of the maximum loss over a short period of time for a predefined probability level. Statistically VaR is the lower quantile on the tail of the portfolio's distribution of possible values at the target horizon. While VaR may still have some usefulness, perhaps as a first-stage tool for risk management, more recent measures of risk provide more reliable tools.

In traditional models of VaR, risk estimates suffer from a number of shortcomings, mainly due to violations of the assumptions on the distributional properties of the underlying risk factors. VaR has been further criticised by those who support the extreme value theory (EVT) like Artzner et al. (1997), Embrechts et al. (1997), Longin (1997) and Embrechts et al. (1998). The main criticism is that VaR estimates do not take into account the magnitude of extreme or rare losses mapped outside the VaR quantile. In fact, VaR considers mainly the frequency of losses, while the severity of a loss is the most important in risk management because a single catastrophic loss could wipe out the firm. Other severe criticisms come from Artzner et al. (1997, 1999), who argue that VaR is not a coherent measure of risk. Using a non-coherent measure such as VaR can result in the portfolio risk being greater than the sum of the risks of the individual components.

VaR has also been criticised for being too conservative, especially during unusual market movements, Barone-Adesi et al. (1998). Since the introduction of VaR, early 1990's, a number of alternative estimation methodologies have been proposed to (a) overcome the limitations of standard methodologies, e.g. Monte Carlo, and (b) match, more closely, the distributional properties of the underlying risk factors, e.g. historical simulation. Nevertheless, whatever the method is, VaR is always defined as the maximum possible loss at a given critical level. Therefore, VaR does not consider any rare, but still possible, loss that is much larger than VaR itself. Those infrequent and unusually large losses can bring a company to bankruptcy.

Filtered historical simulation (FHS) has emerged as one of the best tools for calculating VaR, Zenti and Pallotta (2001). However, financial institutions face problems due to the fact that VaR is not a coherent measure of risk. Expected

---

1 The full variance covariance models like the exponential smoothing and also the (unconditional) historical simulation type models.
2 For a detailed discussion on the shortcomings of the variance–covariance based models, see Giannopoulos, 2000. For a discussion about the shortcomings of the historical simulation, see Pritsker (1999).
3 As Neftci (2000, p. 1) points out “volatility refers to the variance of a random variable, while extremes are a characteristic of the tails only”. And VaR is a scaled measure of risk based on the volatility of the portfolio’s returns (the random variable).
4 For a review of the critics and those who criticised VaR, see Szegö (2002).
5 This is due to the violation of the sub-additivity axiom, see Artzner et al. (1999, p. 209).
6 This is only one of the criticisms of VaR. For a detailed review of all criticisms and a list of those who criticised VaR, see Szegö (2002).
Shortfall (ES) offers a robust theoretical alternative as a measure of risk, although no study has considered in detail the links between ES and FHS.

In this study we show how the FHS can be used in estimating the expected shortfall. The methodology proposed here is quite flexible in a sense that it can handle individual securities and portfolio of securities, provided that a robust conditional volatility model is in place. Since ES is a new risk measure, the main question is whether it is a proper risk measure. Since the class of spectral risk measures contains an infinite number of coherent risk measures, it raises the question whether ES under FHS is a spectral risk measure or comes close to it. We show that the estimator of ES as calculated by FHS is a coherent measure. Moreover, it transpires that this estimator converges for large samples to a spectral risk measure. Therefore we can apply some very general results to develop further analytical tools in this context. Furthermore, the detailed empirical analysis emphasizes the applicability of FHS as an estimation procedure for ES. Finally, we apply two different methods to generate standard errors for the ES estimators. The first method is derived as a closed form asymptotic formula using a general result introduced by Acerbi (2004). The second method employs the formulae derived in Efron (1987) in relation to the accuracy of estimation when using bootstrap repetitions.

The rest of the paper is organised as follows. Section 2 contains an overview of the FHS and the proof that ES as estimated by the FHS is a coherent risk measure. In Section 3 we first describe how to build a FHS tool for estimating the ES, then show why this new measure can be seen as a coherent risk measure and finally we identify the formula for the variance of the ES under FHS. An empirical comparative simulation exercise is detailed in Section 4. Section 5 concludes the paper.

2. Coherent measures of risk

The theory of coherent risk measures was initiated by Artzner et al. (1997, 1999) and developed further by Bertsimas et al. (2000), Delbaen (2000), Kusuoka (2001), Acerbi (2002), Fritelli and Rosazza Gianin (2002), Acerbi and Tasche (2002a), Szegö (2002) and Inoue (2003). Pelessoni and Vicig (2001) defined coherent measures of risk over an arbitrary set of risks, a more general case than the concept of coherence for risk measures over a linear space of random numbers. VaR and the coherent measures mentioned below are static measures of risk. Dynamic risk measures have been proposed by Cvitanic and Karatzas (1998) and Wang (1999), however they seem to be difficult to implement and therefore are not discussed further here.

A risk measure $\rho(\cdot)$ is called coherent if it satisfies the following conditions:

- $\rho(mX) = m\rho(X)$ (homogeneity),
- $\rho(X) \geq \rho(Y)$, if $X \preceq Y$ (monotonicity),
- $\rho(X + Y) \leq \rho(X) + \rho(Y)$ (sub-additivity),
- $\rho(X + a) = \rho(X) - a$ (risk-free condition).
Taking \( m = 0 \) in the homogeneity condition it follows that \( \rho (0) = 0 \) and therefore substituting \( X = 0 \) in the monotonicity condition we get that \( \rho (Y) \leq 0 \), for any \( Y \geq 0 \), as expected. The homogeneity and monotonicity axioms imply that a coherent risk measure is automatically a convex function. Pelessoni and Vicig (2003) generalise the concept of convex risk measures using convex imprecise previsions.

A number of alternative risk measures have been proposed that are simultaneously coherent and may also consider losses beyond VaR. Acerbi and Tasche (2002a) describe and provide mathematical definitions for five measures of risk that include losses in excess of VaR:

- Conditional VaR (CVaR), Rockafellar and Uryasev (2002).
- Expected shortfall (ES), Acerbi and Tasche (2002a).
- Tail conditional expectation (TCE), Artzner et al. (1999).
- Worst conditional expectation (WCE), Artzner et al. (1999).
- Spectral risk measures.

In addition, Testurini and Uryasev (2000) provide one more measure of risk that is called the “expected regret”.

In this study we focus our attention on the ES and we propose for practical implementation an estimator based on FHS, which was introduced by Barone-Adesi et al. (1998) and Barone-Adesi et al. (1999). Hence, we limit our study to the ES, defined here for a critical level \( \alpha \) and an asset \( X \) as

\[
\text{ES}^{(\alpha)}(X) = -E\{X|X \leq \text{VaR}_\alpha\}. \tag{1}
\]

While a more rigorous definition of expected shortfall is provided and its various properties discussed from a theoretical point of view in Acerbi and Tasche (2002a, b), we prefer to use the above tail conditional expectation format, used also by Artzner et al. (1997). The reason is that the two measures are the same over the class of continuous distributions and we stand for a more pragmatic view that only continuous probability distributions are used in practice.

Acerbi and Tasche (2002a, b) provide a proof of the coherence of the expected shortfall together with some other interesting theoretical properties. Expected shortfall was introduced in an equivalent form by Rockafellar and Uryasev (2002) under the name of Conditional Value-at-Risk (CVAR). Most importantly, Inui and Kijima (2003) proved that ES gives the minimum value among a class of plausible coherent risk measures. Moreover, they showed that any coherent risk measure is given by a convex combination of expected shortfalls.

In addition to the theoretical superiority of ES over the VaR, such as the coherence properties, there are further practical issues that recommend once again expected shortfall as a robust risk measure. For example, Yamai and Yoshina (2002a, b) compared the two risk measures in terms of estimation errors, decomposition into risk factors and optimization. They also investigated their validity under market stress. The ES can be easily decomposed and optimized while VaR could not. A small price to pay is that, under some general distributional assumptions, ES requires a larger sample size than VaR at the same level of accuracy. Both measures
seem to underestimate the risk of securities with fat-tailed properties and a high potential for large losses. However, this problem is less acute for ES.

ES has been successfully used for mean-risk optimization portfolio problems. Not only this risk measure has more attractive properties but, as emphasized in Ogryczak and Ruszczyński (2002), it leads to LP solvable portfolio optimization models in the case of returns defined by realizations of discrete random variables under specified scenarios. Schulmerich and Trautmann (2003) showed that only linear programming is needed for hedging a European derivative by developing a self-financing strategy that minimizes the ES.

3. Estimating the expected shortfall with filtered historical simulation

Barone-Adesi et al. (1998, 1999) introduce the FHS algorithm in order to generate correlated pathways for a set of risky assets. At each simulation trial, a value for each asset is generated and all securities in the portfolio are re-priced. After running a large number of simulation trials, a set of portfolio values is generated that forms the empirical distribution for the predicted portfolio values at a certain horizon. The core of FHS is the re-sampling of past returns. This is a non-parametric risk measurement methodology, but unlike other non-parametric methodologies, the past (historical) residual returns are first adjusted by that period’s market conditions, i.e. by scaling them with the corresponding conditional volatility. Thereafter, these standardised residuals are scaled by a volatility forecast that reflects current and future market conditions to form a set of innovations in the multi-period simulation.7

With FHS, the distributional assumptions of the underlying risk factors are relaxed but the current market conditions, i.e. the conditional levels of volatility, are taken into account. Barone-Adesi et al. (1999, 2002) and Barone-Adesi and Giannopoulos (2001) employ the FHS to estimate the market risk on portfolios consisting of linear and non-linear securities. Nevertheless, the main use of FHS in the various studies until now has been focusing on the estimation of market risk, Barone-Adesi et al. (1999, 2002), Barone-Adesi and Giannopoulos (2001), Pritsker (1999), Zenti and Pallotta (2001). Although all these studies found the risk estimates produced by FHS to be more accurate than those predicted by either the historical simulation or parametric methods, no attention has been paid to the fact that FHS can be also employed in the estimation of a number of alternative risk measures such as ES.

The recent methodologies on measuring risk in the lower tail of the distribution of returns have been dominated by methods based on EVT. However, EVT is not free from limitations and criticisms, Embrechts (2001). In contrast to EVT, the FHS can be applied across an unlimited number of risk factors. FHS makes minimal assumptions about the risk factors and still takes into account the current market conditions. With FHS the portfolio is re-priced at each simulation trial and horizon.

7 The “raw” returns, however, are unsuitable for historical simulation because they do not fulfil the properties necessary for reliable results.
That is, FHS takes into consideration positions on contracts that expire soon after the first VaR day, which may cause large shifts in portfolio expected value and could enlarge the portfolio risk. Likewise, all derivative contracts are fully re-priced at each simulation node. The risk measures produced by the FHS are found on the empirical distribution of realistic future portfolio values. Furthermore, FHS takes into account conditional varying volatilities.

Nothing has been said in the literature about the applicability of the FHS to the calculations of coherent measures such as ES. In this study, we show how FHS can be adapted to estimate ES and we discuss whether the estimator obtained in this manner satisfies the sub-additivity condition. There are four main steps in implementing FHS that are summarised below for convenience:

In a first step, a conditional volatility model is fitted to historical data. There are many conditional volatility models in the literature. The right model should produce i.i.d. residuals as well as having a good forecasting power in predicting volatility over the VaR horizon.

The resulted model from step one is to be fitted on the historical data to generate volatilities for each day of the sample period. The realised returns are then standardised by dividing each one of them by the corresponding volatility and the standardised values should be i.i.d. This procedure forms the second step.

The third step consists of bootstrapping from the above sample set of standardised returns a large number \( L \) of drawings. \(^8\) Thereafter each drawing is multiplied by the volatility forecast at the horizon to obtain a sample of i.i.d. values, as large as needed, from the possible profits and losses (P/L) that will occur on that day. Those simulated P/L innovations will form the lagged error term in the next period’s variance equation. \(^9\)

Finally, having a sample of end-of-horizon asset values we can now calculate any statistic we fancy. The FHS procedure can be either used to estimate the ES as the average of a set of VaRs obtained for different confidence levels, Dowd (2002); or directly calculate the estimator as the mean of the truncated sample where the cut-off point is again VaR at \( z \). Once a sample of P/L values at the horizon is obtained by FHS, it is easy to estimate any coherent measure of risk that is described by a closed formula.

The problem investigated here is related to the estimation of such a measure. Although it is widely known that ES is coherent, there are many estimators that can be proposed for the same measure, leading to different estimates. While this is more of a statistical problem it is still very important. ES may be widely accepted as a measure of risk due to its coherence but it is obvious from the formula above that, practically speaking, a bad estimation of VaR implies a miss-estimation of ES. For example, suppose that the estimated VaR is 10% larger than the hypothetical real one. Then, the tail or subsample over which the ES is estimated becomes shorter and, with less values, the estimated mean is more likely to be off the mark.

---

\(^8\) Note that the drawing is made with replacement from a finite set of possible values which are already iid.

\(^9\) The algorithm is described in more details in Section 4.
too. Nonetheless, Acerbi (2004) remarks that ES is less sensitive to the choice of confidence intervals than the VaR. This is obvious since the former measure of risk takes into consideration the whole tail. As such, this robustness property is an extra argument for using ES.

Research on risk measures ought to have a practical linkage and philosophy. The estimation process of ES, and indeed VaR as an intermediate step, is crucial. Our target is to estimate ES. One approach suggested by Dowd (2002) is to slice the tail of the distribution of profits and losses by estimating higher order VaR’s and then averaging those VaR values. This methodology has the advantage of allowing for the best available methods in the estimation of VaR. Inui and Kijima (2003) proposed an extrapolation method to estimate ES.

A more complex procedure for VaR and ES has been proposed by Martins-Filho and Yao (2002), who combine a local polynomial estimator of conditional mean and volatility functions in a conditional heteroskedastic autoregressive nonlinear model with EVT. For finite samples this method seems to outperform estimators currently available in the literature, including that proposed by McNeil and Frey (2000). A multivariate Markov switching process has been used in Guidolin and Timmermann (2003) to model the joint distribution of stock and bond returns and, furthermore, to calculate VaR and ES. Scaillet (2004) considers a non-parametric method that gives a kernel estimator of ES.

An interesting approach has been developed by Merino and Nyfeler (2004) in the context of credit risk modelling. ES is successfully employed to break down the risk of a portfolio into individual counterparties levels, allowing calibration to the financial institutions preferences by choosing an appropriate threshold of losses. The downside of this approach is that Monte Carlo simulation is needed for calculations of ES which can be prohibitive for large portfolios. Merino and Nyfeler (2004) overcame this difficulty by implementing importance sampling, in a conditional independence framework, as a computational mechanism that dramatically increases the rate of convergence of the simulation. Another possibility advocated by Frey and McNeil (2002) is based on Bernoulli mixture models. They showed that the tail of the portfolio loss distribution is represented by the mixing distribution in a Bernoulli mixture representation and in order to calculate ES all what is needed is to calculate the conditional tail expectation of the mixture distribution followed by some appropriate scaling.

However, these approaches lack parsimony and may add very few benefits in contrast to the excess computational efforts that are involved. A more direct methodology makes use of order statistics, as in Acerbi and Tasche (2002b), and was also advocated by Dowd (2001) for VaR calculations. Suppose that a sample $X_1, X_2, \ldots, X_n$ of potential P/L values of the variable $X$ is available and ES is calculated for $\alpha = A\%$. Then, the sample is ordered in an increasing order $X_{[1]} \leq X_{[2]} \leq \cdots \leq X_{[n]}$ and truncated by the VaR value. The number of elements that we need to retain for further calculations is $\eta = \lfloor nA \rfloor$ where $\lfloor \cdot \rfloor$ is the integer part function. The worst $\alpha = A\%$ losses are then $X_{[1]} \leq X_{[2]} \leq \cdots \leq X_{[\eta]}$ and the obvious ES estimator is the average of the highest $A\%$ losses from the set of $X_1, X_2, \ldots, X_n$. 
Acerbi and Tasche (2002b) showed that

\[
\overline{\text{ES}}^{(2)}(X + Y) = -\frac{1}{n} \sum_{i=1}^{n} (X + Y)_{[i]} \leq -\frac{1}{n} \left\{ \sum_{i=1}^{n} X_{[i]} + \sum_{i=1}^{n} Y_{[i]} \right\} 
= \overline{\text{ES}}^{(2)}(X) + \overline{\text{ES}}^{(2)}(Y)
\]  

(3)

Since FHS proved to be a very reliable estimation procedures for VaR, as shown by Barone-Adesi et al. (2002), and Zenti and Pallotta (2001), it seems natural to extend its application to the calculation of ES. It is important to know how a newly proposed risk management measure can be calculated for individual securities as well as at the portfolio level. Without loss of generality, consider a portfolio made of two assets \(X\) and \(Y\). Having a historical sample \(((X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n))\) we need the conditional volatilities for each asset and for the portfolio of the two assets. FHS scales the above sample by dividing by the corresponding volatilities and thus produces two iid samples:

\[
\tilde{X}_i = \frac{X_i}{\sigma_{x,t-i}}, \quad \tilde{Y}_i = \frac{Y_i}{\sigma_{y,t-i}}, \quad i = 1, \ldots, n.
\]  

(4)

Each sample is re-scaled back using the corresponding forecast of volatility \((\sigma_x, \sigma_y)\) at the end of the horizon. When bootstrapping with replacement from the sample

\[
\left((\sigma_x \tilde{X}_1, \sigma_y \tilde{Y}_1), (\sigma_x \tilde{X}_2, \sigma_y \tilde{Y}_2), \ldots, (\sigma_x \tilde{X}_n, \sigma_y \tilde{Y}_n)\right)
\]

we get another sample \(((\sigma_x \tilde{X}_1, \sigma_y \tilde{Y}_1), (\sigma_x \tilde{X}_2, \sigma_y \tilde{Y}_2), \ldots, (\sigma_x \tilde{X}_n, \sigma_y \tilde{Y}_n))\) of any size \(L\) that is needed which can be used for the calculation of any statistic and in particular ES. It is important to realise that, due to the bootstrapping procedure which draws with replacement, \(\tilde{X}_j\) can take any of the values \(\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_n\), for \(j = 1, \ldots, L\) and similarly \(\tilde{Y}_j\) can take any of the values \(\tilde{Y}_1, \tilde{Y}_2, \ldots, \tilde{Y}_n\), for \(j = 1, \ldots, L\). As shown above, at the critical level \(\alpha\), we can estimate the ES of each asset at the univariate level,

\[
\overline{\text{ES}}^{(2)}_{\text{FHS}}(X) = -\frac{\sigma_x}{\theta} \sum_{i=1}^{\theta} \tilde{X}_{[i]}, \quad \overline{\text{ES}}^{(2)}_{\text{FHS}}(Y) = -\frac{\sigma_y}{\theta} \sum_{i=1}^{\theta} \tilde{Y}_{[i]},
\]  

(5)

where \(\theta = \lfloor L\alpha \rfloor\). As discussed by Acerbi and Tasche (2002a), we should be careful when calculating the ES when \(L\alpha\) is not a positive integer. However, in the context proposed here, since one may bootstrap as many observations as needed, this problem is overcome by taking \(L\) large enough to ensure that \(L\alpha\) is a positive integer, in which case \(\theta = L\alpha\).

---

10 Here we have simplified slightly the notation by leaving out the time index. However, if the horizon is longer than one day FHS is based on simulating entire paths until a value at the horizon is obtained.
For the same critical level we calculate the ES for the portfolio of the two assets $X + Y$ rescaling the sample of pairwise sums by the volatility $\sigma_{X+Y}$ of $X + Y$. \(^{11}\)

Now we show an important result about the ES estimated by the FHS. Following Acerbi (2004), we prove next that this risk measure calculated as described above is a coherent risk measure. Recall first that if $U_1, U_2, \ldots, U_N$ are $N$ realisations of the portfolio profit-loss $U$ then for any $N$-tuple of weights $\{\phi_i\}_{i=1}^{N}$ the statistic

$$M^{(N)}_\phi(U) = - \sum_{i=1}^{N} U_{[i]} \phi_i$$

is coherent for any given positive integer $N$ if and only if the following three properties are simultaneously satisfied:

(i) $\phi_i \geq 0$,
(ii) $\phi_j \leq \phi_i$ if $i < j$,
(iii) $\sum_{i=1}^{N} \phi_i = 1$.

An $N$-tuple of weights $\{\phi_i\}_{i=1}^{N}$ satisfying those three conditions is called an admissible risk spectrum, Acerbi (2004).

Since $(\sigma_x \tilde{X}_1, \sigma_x \tilde{X}_2, \ldots, \sigma_x \tilde{X}_L)$ is the sample of size $L$ used to calculate ES under FHS, \(^{12}\) it is evident that $L = N$ and $U_i = \sigma_x \tilde{X}_i$ for any $i = 1, \ldots, L$. Then, using formula (5) it follows that:

$$\overline{\text{ES}}^{(x)}_{\text{FHS}}(X) = - \frac{\sigma_x}{\theta} \sum_{i=1}^{\theta} \tilde{X}_{[i]} = - \sum_{i=1}^{\theta} \frac{1}{\theta} \sigma_x \tilde{X}_{[i]} = - \sum_{i=1}^{\theta} \frac{1}{\theta} \sigma_x \tilde{X}_{[i]} - \sum_{i=\theta+1}^{L} 0 \times \sigma_x \tilde{X}_{[i]}.$$

Therefore the admissible risk spectrum in this case is given by

$$\phi_i = \frac{1}{\theta}, \quad \text{for } i = 1, \ldots, \theta \quad \text{and} \quad \phi_i = 0 \quad \text{for } i = \theta + 1, \ldots, L. \quad (7)$$

It is obvious that it satisfies the conditions (i)–(iii) above. Thus $\overline{\text{ES}}^{(x)}_{\text{FHS}}(X)$ is a coherent measure for any number of drawings $L$.

An important issue for backtesting procedures and auditing in general is the estimation error of the risk measure used. It is known that the VaR has a lower statistical error than ES, but this is not surprising since the latter measure is based on all observations in the tail, not just one.

In order to determine a formula for the variance of the risk measure we use a very general result \(^{13}\) that gives the variance of a spectral measure. For a risk spectrum

\(^{11}\) This is obviously $\sigma_{X+Y}^2 = \sigma_x^2 + \sigma_y^2 + 2\sigma_{xy}$ where all three terms on the right side are varying in time.

\(^{12}\) For one day horizon we have the same conditional volatility $\sigma_x$ that is scaled in the standardised residual returns. For a horizon of 10 days period, different paths lead to different conditional volatilities at the end of horizon. The methodology detailed here for the one-day horizon can be easily adapted so that the results derived here are still valid. For a 10 days horizon one may take $U_i = \tilde{X}_i$ where the term on the right side already includes conditional volatilities at the end of the path.

\(^{13}\) This is proposition 5.1. from Acerbi (2004).
\( \phi: p \to \phi(p) \) and a profit and loss variable \( U \) with a known cdf \( F \) of the variance of the spectral measure \( M^{(N)}(U) \) is

\[
\sigma^2 = \frac{2}{N} \int_{u<v} du \, \phi(F(u)) \phi(F(v)) F(u)(1 - F(v)).
\]

(8)

The risk spectrum

\[
\phi(p) = \begin{cases} 
\frac{i}{n} & \text{if } p \leq \frac{\theta - 1/2}{L}, \\
0 & \text{if } \frac{\theta - 1/2}{L} < p,
\end{cases}
\]

(9)

gives exactly the admissible spectrum in (7) for the mapping \( \phi_i = \frac{1}{L} \phi(i-1/2) \). In order to calculate (8) we need to specify the cdf \( F \). The bootstrapping is done with replacement where all historical simulations have equal chances to be drawn. Therefore, with a historical sample of size \( n \), the cdf \( F \) is that of a discrete distribution

\[
F(u) = \begin{cases} 
0, & \text{if } -\infty < u < \sigma_x X_1, \\
\frac{i}{n}, & \text{if } \sigma_x X_1 \leq u < \sigma_x X_2, \\
\frac{2}{n}, & \text{if } \sigma_x X_2 \leq u < \sigma_x X_3, \\
& \vdots \\
1, & \text{if } \sigma_x X_n \leq u < \infty.
\end{cases}
\]

(10)

For notational simplicity \(-\infty \) denotes to \( \sigma_x X_0 \) and \( \infty \) denotes to \( \sigma_x X_{n+1} \). From (8) with \( N \) replaced by \( L \) and using the cdf in (10) it follows that:

\[
\sigma^2 = \frac{2}{L} \int_{u<v} du \, \phi(F(u)) \phi(F(v)) F(u)(1 - F(v))
\]

\[
= \frac{2}{L} \sum_{i<j} \int_{A_i \cap A_j} \phi\left( \frac{i}{n} \right) \phi\left( \frac{j}{n} \right) \frac{i}{n} \left( 1 - \frac{j}{n} \right) du \, dv
\]

(11)

where \( A_i = [\sigma_x X_i, \sigma_x X_{i+1}] \) for any \( i = 0, 1, \ldots, n \). In order to calculate the above integrals it is crucial to know whether \( \frac{i}{n} \leq \frac{\theta - 1/2}{L} \) which is equivalent \(^{14}\) to \( \frac{i}{n} \leq \alpha - \frac{1}{2L} \). Consider \( \hat{i}^{(n)}_s \) the largest positive integer \(^{15}\) such that

\[
\frac{\hat{i}^{(n)}_s}{n} \leq \alpha - \frac{1}{2L}.
\]

\(^{14}\) Recall that \( L \) is large enough to have \( \theta = L \alpha \).

\(^{15}\) From a theoretical point of view it is possible to construct a situation where \( \hat{i}^{(n)}_s \) is 0. This can happen when \( \alpha \) is very small and \( n \) is relatively small. Since \( L \) can be made arbitrarily large we only need to compare \( \hat{i}^{(n)}_s \) and \( \alpha \). At 1% critical level, for any \( n \) less than 100 it is obvious that \( \hat{i}^{(n)}_s \) is 0. However the formula for the variance of the estimator of risk measure is true asymptotically and the assumed advantage of employing the filtered historical simulation is that sufficient historical data are available.
Then

\[\sigma^2 = \frac{2}{L} \sum_{0 \leq i < j \leq L} \phi \left( \frac{j}{n} \right) \phi \left( \frac{i}{n} \right) \left( 1 - \frac{j}{n} \right) \sigma_x^2 (X_{i+1} - X_i)(X_{j+1} - X_j)\]

\[\sigma^2 = \frac{2\sigma_x^2}{Ln^2} \sum_{0 \leq i < j \leq L} (n-j)(X_{i+1} - X_i)(X_{j+1} - X_j)\]

(13)

4. An empirical investigation

We will illustrate our methodology with the use of a numerical example. Our dataset consists of daily returns from the NASDAQ 100 index covering the period of 2nd January 1998 to the 31st December 2003, which includes a total of 1507 observations. We chose this index because it is characterised by high volatility and, as we will see below, by excess kurtosis. These distributional properties imply large losses which is a characteristic that can challenge a risk model.

We tested for peakedness, fat tails and leptokurtosis using the Schmid and Trede (2004) test. The test is based on the selector statistics of the quantiles. It checks for peakedness and fat tails separately. It does not make any assumptions about the distribution that describes the data and it is not very sensitive to outliers. The test statistics with the standard errors are reported on Table 1. \(T_n, P_n\) and \(L_n\) are the test statistics for tailness, peakedness and leptokurtosis respectively.

Table 1 shows that the test are significant at the 99% probability level for all three hypotheses tested. Furthermore, we found that the series of returns is dominated by conditional heteroskedasticity.\(^{16}\) A Garch type of family models was fitted to the data. We based our conditional volatility model specification on two main criteria; (i) the model must produce standardised residuals that are i.i.d.\(^{17}\) and (ii) have the highest and unbiased forecasting power. We reached the following model specification:

\[r_t = \theta \varepsilon_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, h_t),\]  

(14a)

\[h_t = \omega + \alpha (\varepsilon_{t-1} + \xi)^2 + \beta h_{t-1}.\]  

(14b)

The MA term, \(\theta\), is to remove the serial correlation that may be present in the return series. The variance in Eq. (14b) is specified as a GARCH(1,1) but an additional term, is imposed in order to capture any asymmetries in volatility that may govern the series. The coefficient estimates and their standard errors are reported on Table 2.

Diagnostic tests confirmed that residuals are random without any dependency in the first or second moments. We tested the ability of the model in Eq. (14) to predict the conditional variance of the innovation term, by regressing the squared innovations at period \(t\) on the conditional variance, \(h_t\). The \(R^2\) coefficient of the regression

\(^{16}\) We conducted several tests for heteroskedasticity on the return series and all revealed presence of strong GARCH effects.

\(^{17}\) Randomness is an essential property for any sample used as innovations in any simulation trial.
was 30%, indicating that the model can predict next day’s variance with satisfactory accuracy. We also found that this forecast is free from any systematic bias.

The standardised residual returns obtained from the GARCH model in (14), $z = \hat{e}_t / \sqrt{\hat{h}_t}$, form the set of i.i.d. innovations in the FHS. A large set of simulated pathways for the conditional mean and variance equations can be formed using simulated innovations, $e^*$, in (14a) and (14b), where

$$e = z \sqrt{\hat{h}_t}.$$  

Thus, a simulated innovation is created by a randomly drawn (with replacement) historical standardised residual return, $z$, which has been rescaled by the last period’s conditional volatility $\sqrt{\hat{h}_{t-1}}$ in order to reflect current market conditions.  

The appealing feature of the FHS algorithm is that it uses innovations from the empirical distribution of the data, i.e. it does not impose any assumptions on the behaviour of the risk factors. However, unlike other non-parametric techniques, e.g. historical simulation and weighted historical simulation, FHS uses i.i.d. standardised residuals, $e$, as innovations which, as they have been re-scaled with the current volatility, reflect current market conditions.

The histogram of 5000 simulated scenarios for the portfolio that corresponds to 1 and 10 days horizons are shown in Fig. 1. Since this is a linear portfolio, the empirical density for day one matches the shape of the density of the historical standardised returns.  

The set of profits and losses in each horizon is computed as $P_{T+N} - P_T$, where $P_T = 1467.92$ is the value of the Portfolio on the last trading day. On a long portfolio the VaR at 99% and 95% probability is measured by the 1st and 5th percentiles. Likewise, the 99th and 95th percentiles coincide to a VaR at 99% and 95% for a short

---

Table 1
Schmid and Trede test for fat tails, peakedness and leptokurtosis

<table>
<thead>
<tr>
<th>Tn</th>
<th>Pn</th>
<th>Ln</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.80363**</td>
<td>1.85686**</td>
<td>3.34908**</td>
</tr>
</tbody>
</table>

Table 2
Coefficient estimates for the GARCH in (14)

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>$\theta$</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\xi$</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimate</td>
<td>-0.0681</td>
<td>4.0691</td>
<td>0.0729</td>
<td>0.9103</td>
<td>-10.3224</td>
<td>-7113</td>
</tr>
<tr>
<td>Standard error</td>
<td>0.026</td>
<td>2.44</td>
<td>0.021</td>
<td>0.031</td>
<td>2.905</td>
<td></td>
</tr>
</tbody>
</table>

---

18 For a detailed description of the Filtered Historical Simulation algorithm, see As Barone-Adesi et al. (1998). For an application to a multi-risk factor model, see Barone-Adesi et al. (1999).

19 For the first day you do not have a proper path, i.e. the conditional volatility in (14b) is deterministic. In effect it is the same as taking values from the density of historical data. The only minor point is that you need to bootstrap enough observations to get a better shape of the density.
Selected percentiles for the portfolio simulated profits and losses for 1 and 10 days horizons are shown on Table 3.

Following (1) the ES is computed as the average of the losses that exceed VaR. Table 4 shows the ES estimates for both long and short positions at a probability level of 99% and 95%.

As with any statistical estimators, there is an estimation error that needs to be taken into consideration. The variance or standard deviation of the newly proposed ES estimator can be a very informative tool. With FHS, both VaR and ES are estimated jointly based on the same sample. Nevertheless, both VaR and ES are based on a simulation solution rather than an analytical one. The density function that describes the risk factors is approximated using the scaled innovations on the risk factor itself to update its conditional mean and variance equations in (14). Furthermore, the ES under the FHS will only converge to a spectral risk measure if the sample data of the exceedances is very large. To get a coherent ES estimates it will be necessary to conduct a very large number of simulation trials (simulated portfolio pathways).

However, generating a large number of simulation trials is not an option, obtaining standard errors of the VaR and ES estimates will provide an insight on the accuracy of these risk parameters. In this study we estimated standard errors using a non-parametric bootstrapping method, as suggested by Efron (1987) and Efron and Tibshirani (1993). They generate multiple new samples from the data sample.

---

20 The portfolio value at the last day of 2002 was 1467.92.

21 In our case the size of the data-sample, of the exceedances, is known as it is a direct function of the number of trials in the bootstrap simulation and the probability level used in defining the VaR.
and calculate the value of the estimator $\hat{\theta}$ in each sample. Let us denote the value of $\theta$ in each FHS sample as $\theta^*$, then

$$\text{Var}_F(\hat{\theta}) \approx \text{Var}_F(\theta^*)$$

where $F$ is the unknown distribution of the estimator $\hat{\theta}$. Hence, we estimate the unknown density $F$ of our ES estimates by repeating the FHS several times. At each simulation we obtain an estimator $\text{ES}^*$ of the expected shortfall as defined in (1). The variance and the percentiles of this estimator are computed from the empirical density $F$ of the above sample. The level of accuracy obtained with the above technique increases with the number of bootstrap repetitions, $B$. The choice of $B$ is particularly important in our case since the sample of exceedances utilised in obtaining a single ES estimate is a small fraction of the number of draws, $L$, in a single FHS. The variance of our estimator can be reduced at a computational cost. There is a trade-off between computational efforts and time versus accuracy. This problem has two dimensions; the number of draws, $L$, in each FHS by the number of FHS repetitions, $B$, employed to obtain the sample of $\text{ES}^*$. The accuracy can be defined as the percentage deviation of the $\widehat{\text{ES}}$ from the ideal ES, obtained when $B = \infty$. Andrews and Buchinsky (1997) suggested a three steps process in order to determine the correct number of repetitions, $B$. After defining a bound, $pdb$, on what will be an acceptable percentage deviation of the $\widehat{\text{ES}}_B$ and the population $\text{ES}_\infty$, in the first set, a threshold of FHS repetitions is computed as follows:

$$B_0 = \text{int} \left( \frac{5000 \gamma^2_{1-\tau}(2 + \gamma)}{pdb^2} \right). \quad (16)$$

The term $1 - \tau$ stands for the probability level of acceptance in the $\chi^2$ statistic and the term in bracket is the degrees of freedom. Thus, the number of repetitions, $B$, is a trade-off between the confidence level in the $\chi^2$ and the level of acceptance of error in the sample $\widehat{\text{ES}}_B$ from the population $\text{ES}_\infty$. As the number of bootstrap repetitions increases the level of accuracy $pdb$ grows at an exponential rate. The term $\gamma$ is the excess kurtosis of the sample of FHS repetitions, which initially is set to zero.

---

Table 3
VaR estimates for day 1 and 10

<table>
<thead>
<tr>
<th></th>
<th>99% Long</th>
<th>95% Long</th>
<th>99% Short</th>
<th>95% Short</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>60.45</td>
<td>47.97</td>
<td>61.52</td>
<td>48.01</td>
</tr>
<tr>
<td>Day 10</td>
<td>221.54</td>
<td>153.51</td>
<td>222.35</td>
<td>155.49</td>
</tr>
</tbody>
</table>

Table 4
ES estimates for day 1 and 10

<table>
<thead>
<tr>
<th></th>
<th>Long 99%</th>
<th>Long 95%</th>
<th>Short 99%</th>
<th>Short 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>80.05</td>
<td>61.51</td>
<td>76.77</td>
<td>59.79</td>
</tr>
<tr>
<td>Day 10</td>
<td>278.99</td>
<td>201.09</td>
<td>262.64</td>
<td>199.13</td>
</tr>
</tbody>
</table>
After choosing a level of error \( pdb \) equal to 10\% and a confidence level \((1 - t)\) equal to 95\%, the initial value of bootstrap repetitions, \( B_0 \), was found to be equal to 200.

The second step consists for running the FHS \( B_0 \) times and obtaining a sample of ES estimates, \( \hat{E}S_{B_0} \). As in Barone-Adesi et al. (1999) we chose the number of simulation trials (scenarios) in each FHS repetition to be 5000. Thereafter, we estimated, in each FHS repetition, the VaR\(_{99\%}\) as the first percentile (i.e. the 50th of the ordered scenarios) and the ES as in (1).

The final step consists of computing

\[
B_1 = \text{int} \left( 2500 \frac{\chi^2_{1-t} (2 + \hat{\gamma})}{pdb^2} \right)
\]

where \( \hat{\gamma} \) is the excess kurtosis of the sample of \( B_0 \) repetitions obtained from step one in (16). The number of optimal bootstrap repetitions is given by \( B^* = \max\{B_0, B_1\} \). In our data-sample of 200 ES estimates the excess kurtosis \( \hat{\gamma} \) was found to be negative and so we set \( B^* = B_0 \).

The empirical probability density functions, for horizons of 1 and 10 days, are illustrated in Fig. 2. They suggest that the density is skewed and fat tailed. The central limit theorem permits a better insight into the asymptotics of the ES estimators.

One of the main issues here is the accuracy of the ES estimators which can be controlled up to some degree as discussed above. In Table 5 we report the descriptive statistics for the sample of ES estimators.

As discussed in Section 3 the ES, estimated by the FHS, is a coherent measure and, as an estimator converges consistently towards a spectral risk measure, see

![Fig. 2. Bootstrapped ES estimates.](image-url)
Acerbi (2004) for further clarifications. By estimating the mean and the standard error as well the higher percentiles, we capture the entire spectrum of that risk location. This will give the option to a more conservative investor to select an ES estimator at a higher quartile or percentile.

For comparative purposes we also calculate the standard errors using the closed formula (13). For one day ahead we get a standard deviation of 0.78 for the estimate of ES. This is obviously smaller than the standard deviation of 3.52 reported in Table 5 and this is not surprising since the closed-formula (13) works asymptotically (Acerbi, 2004). This is equivalent to imposing an extremely high level of accuracy, and, of course, the standard deviation of the ES estimate should then shrink dramatically.

5. Conclusions

In this study we have showed how the FHS can provide coherent risk estimates. It is emphasized that, from the estimation point of view, VaR is still important as determining the cut-off point where the sample of profits and losses values should be truncated for further estimation of more extreme measures of risk. It has been shown above that the ES under FHS is a coherent measure that converges for large samples to a spectral risk measure. A general variance formula for this type of risk measures was applied to determine the possible statistical error made by using this methodology.

ES under FHS combines one of the best theoretical risk measures with one of the best applied econometrical modelling techniques used in risk management. The results described here give hope for a wider applications of this methodology for measuring and monitoring market risk.

An empirical exercise using the daily returns for the NASDAQ 100 shares index between 1998 and 2003 revealed that the longer the horizon estimates of ES and VaR, as computed by the FHS, are less precise. The results obtained here confirm earlier evidence in the literature that the ES estimates are more uncertain than the VaR estimates. The accuracy of an estimator of any risk measure is an important issue. In the context of using ES under FHS, we implemented a known methodology for controlling the degree of accuracy of an estimator obtained by bootstrapping. In addition, since we have proved that ES under FHS is a coherent risk measure, we were able to derive a closed form asymptotic formula for the standard deviation

<table>
<thead>
<tr>
<th></th>
<th>ES</th>
<th>VaR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 1</td>
<td>75.11</td>
<td>59.62</td>
</tr>
<tr>
<td>Day 10</td>
<td>258.67</td>
<td>216.13</td>
</tr>
<tr>
<td>Std</td>
<td>3.52</td>
<td>0.86</td>
</tr>
<tr>
<td>1st %ile</td>
<td>82.64</td>
<td>61.43</td>
</tr>
<tr>
<td></td>
<td>276.01</td>
<td>229.36</td>
</tr>
</tbody>
</table>
of this estimator. This can be used as an additional tool for monitoring the volatility of the FHS estimator of ES.

Acknowledgments

We are grateful for the helpful comments and suggestions of Carlo Acerbi, Kilani Ghoudi and Les Vosper.

References

Andrews, D., Buchinsky, M., 1997. On the number of bootstrap repetitions for bootstrap standard errors, confidence intervals, and tests, Cowles Foundation paper 1141R.


