EDHEC RISK AND ASSET
MANAGEMENT RESEARCH CENTRE

393-400 promenade des Anglais
06202 Nice Cedex 3
Tel.: +33 (0)4 93 18 32 53
E-mail: research@edhec-risk.com
Web: www.edhec-risk.com

A Comparison of Alternative Approaches for Determining the Downside Risk of Hedge Fund Strategies

January 2007

Daniel Giamouridis
Lecturer in the Department of Accounting and Finance at the Athens University of Economics and Business, Visiting Fellow in the Faculty of Finance of the Sir John Cass Business School, City University, London and Research Associate with the EDHEC Risk and Asset Management Research Centre

Ioanna Ntoula
Risk Analyst in the Methods and Measures Division, Market Risk Management Sector, Eurobank EFG, Greece

EDHEC
BUSINESS SCHOOL
Lille - Nice

ABSTRACT

This paper compares a number of different approaches for determining the Value at Risk (VaR) and Expected Shortfall (ES) of hedge fund investment strategies. We compute VaR and ES through completely model-free methods, as well as through mean/variance and distribution model-based methods. Among the models considered certain specifications can technically address autocorrelation, asymmetry, fat tails, and time-varying variances which are typical characteristics of hedge fund returns. We find that conditional mean/variance models coupled with appropriate distributional assumptions improve our ability to predict VaR, 1% VaR in particular. We also find that the goodness of ES prediction models is primarily influenced by the distribution model rather than the mean/variance specification.

JEL classification: C16; C52; G10; G23

Keywords: Value at Risk; Expected Shortfall; Hedge Funds; GARCH; Backtesting

Authors' note: This article is a working paper currently under review for possible publication.

Acknowledgements

We are grateful to Vassilios Siokis, Chief Risk Officer at Cheyne Capital Management Limited for many helpful discussions and comments. Marc Paollela and Ioannis Vrontos are also greatly acknowledged for their remarks.

EDHEC is one of the top five business schools in France owing to the high quality of its academic staff (104 permanent lecturers from France and abroad) and its privileged relationship with professionals that the school has been developing since its establishment in 1906. EDHEC Business School has decided to draw on its extensive knowledge of the professional environment and has therefore concentrated its research on themes that satisfy the needs of professionals.

EDHEC pursues an active research policy in the field of finance. Its "Risk and Asset Management Research Centre" carries out numerous research programs in the areas of asset allocation and risk management in both the traditional and alternative investment universes.

Copyright © 2007 EDHEC

1. INTRODUCTION

The hedge fund or alternative investments sector has been the most rapidly growing sector of financial services over the recent years. Many hedge funds yielded significant returns to their investors, even during periods when traditional asset classes performed poorly. The industry, however, has also experienced various investment disasters with the failure of Long-Term Capital Management (LTCM hereafter), claimed to have nearly blown up the world's financial system (Jorion, 2000). As a consequence risk control has received significant attention and an enormous effort has been dedicated in defining universal risk management principles for the hedge fund sector.[1] The growing demand for hedge fund exposure, by high net-worth individuals or institutional investors, but also by small- to medium-sized investors through investable hedge fund indices, or via index-linked products, gives an additional boost to the interest in identifying appropriate risk management tools and highlights the need for more academic and practitioner studies in the area of hedge fund risk measurement and control.

An aggregate measure of risk for traditional investment portfolios is the Value at Risk, VaR, i.e., the worst loss that can happen under normal market conditions over a specified horizon (e.g., 1 day) at a specified confidence level (e.g., 99%). Leverage, derivatives, event arbitrage, and other complex strategies used by managers expose alternative investments to a plethora of risks,[2] and may in some cases make VaR an inappropriate measure of the entire risk exposure of a hedge fund. However, as an established risk measurement and reporting tool in the banking and finance sectors for years, VaR helps to communicate hedge fund risks to institutional investors in a relatively disciplined fashion, and has thus become part of the hedge fund risk managers' daily risk analysis. In addition, it is required in the funds' due diligence questionnaires and is also reported in monthly risk statements distributed to the funds' investors. VaR, finally, receives considerable attention in the vast majority of groups working on the establishment of risk management principles in the hedge fund industry (see footnote 1).

Academic- and practitioner-oriented research in hedge fund risk measurement has, as a result, focused primarily on VaR (e.g., Bali and Gokcan, 2004, Gupta and Liang, 2003, Jeager, 2002, Jorion, 2000, Lhabitant, 2001). Other risk measures that have been considered include variations of VaR, such as the Expected Shortfall, or ES (see, e.g. Agarwal and Naik, 2004, Giamouridis and Vrontos, 2006, Gupta et al., 2005, Krokhmal et al., 2002), the Drawdown, DD, (e.g., Krokhmal, et al., 2002, López de Prado and Peijan, 2004), and the Time Under Water, or TUW (López de Prado and Peijan, 2004). This literature provides many useful insights in terms of hedge fund risk measurement and control. Jorion (2000), studies the VaR of LTCM. He stresses that the LTCM's risk measurement method fails primarily for two reasons, both related to the unique characteristics of the investment strategy, i.e., long positions in credit-sensitive instruments: the assumption of constant volatility and the assumption of symmetric profit/loss distribution. Lhabitant (2001) measures VaR in terms of an alternative measure, the 'Value at Market Risk'. The latter is based on the extreme moves of individual risk factors estimated from the factor return distribution using Extreme Value Theory (EVT hereafter). Analysis of a large dataset of hedge funds monthly returns produces inconclusive results most probably because of the short backtesting period. Gupta and Liang (2003) are concerned with the VaR of hedge funds with an eye to assess the funds' capital adequacy. They use monthly hedge fund return data for individual funds and consider EVT a firm theoretical framework to compute VaR. More recently, Bali and Gokcan (2004), provide a comparison of four different approaches, based on the normal distribution, the generalized error distribution, the generalized Pareto distribution, and the Cornish-Fisher approximation. Using monthly data for hedge fund strategy indices they conclude that the VaR estimations based on the generalized Pareto distribution and the Cornish-Fisher approximation perform best in terms of their in-sample quality-of-fit tests. Another recent work by López de Prado and Peijan (2004) studies DD and TUW of hedge fund strategy indices. It concludes that whenever the profit/loss distribution is not symmetric and/or returns are time-dependent the VaR does not capture all dimensions of market risk. Other studies that have been concerned with the measurement of hedge fund risk include Agarwal and Naik (2004), Giamouridis and Vrontos (2006), Gupta et al. (2005), and Krokhmal et al. (2002). These, however, study the downside risk of hedge funds in the context of portfolio construction. Finally, Lo (2001) discusses comprehensively several important theoretical and practical issues related to risk management for hedge funds.

We measure downside risk of hedge funds in terms of the two most established risk measures, VaR and ES, and compare several common models. These include unconditional mean/variance models, i.e., historical simulation

1 - See, e.g. The President's Working Group on Financial Markets (1999), MFA (2000, 2003, 2005), IRC (2000, 2001), Basel Committee on Banking Supervision and the International Organization of Securities Commissions (2001).

2 - Lo (2001) and Jeager (2002) provide a detailed discussion on the risks associated with hedge fund investments, i.e., market risk, credit risk, liquidity risk, and event risk among others.

with different time windows, as well as conditional mean/variance models combined with filtered historical simulation, and normal, Cornish-Fisher, and generalized Pareto distribution parameterizations. Some of these models exhibit features that can possibly deal with concerns expressed in the literature, i.e., that, 'conditional' measures of risk are more relevant than 'unconditional' measures (Lo, 2001), returns are often serially correlated (Getmansky, et al., 2004), returns exhibit dynamic volatility (Giamouridis and Vrontos, 2006), the conditional distribution of returns is asymmetric and fat-tailed (Bali and Gokcan, 2004, Gupta and Liang, 2003). Our study is aided by the recent launch of Hedge Fund Research (HFR hereafter) daily index return database. In addition to making possible the estimation of the former models, this database allows us to calculate VaR and ES for shorter horizons. The 1-month horizon used in the previous studies may be appropriate to compute the fund's capital requirements,[3] but is not necessarily adequate for communicating the risks to investors with short-term views, i.e., funds of funds, hedge fund index-linked products' investors.

We find that conditional mean/variance models improve our ability to predict VaR, on the basis of producing lower average violations and fewer threshold violations. When coupled with appropriate distributional assumptions, conditional mean/variance models improve, predicting VaR thresholds further in the tails, i.e., 1% VaR. Our analysis for ES is not that conclusive given the amount of available data. We find that in ES prediction what matters more is the distributional assumption. In particular, predictions based on a relatively large history of data —model-free approach—compare equally well with conditional mean/variance models but only with those combined with appropriate distributional assumptions. These results are appealing to managers, multi-strategy or index-linked products managers primarily, wishing to establish an efficient risk measurement and monitoring strategy for a short-term horizon. They also provide useful insights for the managers who construct portfolios on the basis of alternative—to the minimum variance—risk objectives, i.e., minimum VaR or ES.

The contributions of this article are several. To our knowledge, model time-varying means and variances of hedge fund returns for the first time in the context of hedge fund risk measurement. López de Prado and Peijan (2004) is the only study we know that introduces time-dependence in the mean specification and concludes that the results found by static models are misleading. Lo (2001) also highlights that more realistic statistical assumptions such as time-varying risk factors and leptokurtic distributions can improve on the appropriateness of VaR for hedge fund risk management purposes. In addition, we provide a comparative analysis of ES prediction models for hedge funds. Agarwal and Naik (2004), Krokhmal et al. (2002), are concerned with ES in the context of hedge fund portfolio construction without examining potential benefits among different ES models. Giamouridis and Vrontos (2006) touch upon this issue by focusing on various variance/covariance models under a multivariate normal distribution model. Although in a univariate setting, we provide additional evidence for several distribution models. We also provide an analysis of the short-term VaR and ES. Existing studies compute VaR and ES over a 1-month horizon, not always adequate, as already discussed, for communicating the risks to investors with short-term views. Finally, we provide a backtest analysis. While the limitations are acknowledged and the results may only be indicative, an analysis of this kind is also new, to our knowledge, to the hedge fund risk management literature.

The paper is structured as follows. Section 2 develops the methods used in the calculation of VaR and ES. Section 3 illustrates the procedures used in the backtesting analysis. The dataset is detailed in section 4 along with the discussion of the empirical analysis and the results. Section 5 concludes the article.

2. METHODS

Calculation of risk measures for portfolios of assets, hedge funds, for example, may involve a position-level analysis, i.e., analysis of the risk of each asset holding. VaR or any other risk measure is then calculated in a multivariate model setting. The added value of this approach, over the alternative of analysing aggregate return of the portfolio, is questioned by Berkowitz and O'Brien (2002). They find that, in their sample, disaggregating the risk of large and complex positions does not outperform an appropriate univariate forecast of VaR. On the other hand, the proprietary nature of hedge fund positions does not allow a meticulous multivariate analysis

3 - Jorion (2000) stresses that the time horizon of VaR should reflect the liquidity of the assets or the time necessary for an orderly liquidation (which depends on the size of the funds). Alternatively, the horizon should cover the time necessary to raise additional funds or for corrective action.

from non-insiders, i.e., fund of funds managers, given that the available information, normally obtained through the due diligence questionnaires, is not very informative on the asset holdings. Finally, even in cases where a multivariate approach is required for certain reasons and, of course, is possible, establishing an appropriate univariate specification could help in deciding which multivariate models are worth pursuing (Kuester et al., 2006). For these reasons, we focus on the univariate approach.

There are currently numerous approaches for predicting VaR and ES. A recent work by Kuester et al. (2006) provides an excellent overview. We focus on models primarily motivated by the nature of hedge fund investments but also on 'naïve' models that have been shown to perform well in practice. This section provides the details and the reasoning underlying our choice of the methods for hedge fund VaR and ES predictions. We define VaR and ES for time *T+1* with coverage probability p (typically p = 0.01 or 0.05), based on information at time *T*, as

$$Pr\left(L_{T+1} > VaR_{t+1}^{p} \middle/ \Omega^{T}\right) = p$$

$$ES_{T+1}^{p} = E\left(L_{T+1} \middle/ L_{T+1} > VaR_{t+1}^{p}, \Omega^{T}\right)$$

respectively, where Ω^{T} denotes the information set known at time T (as in Christoffersen and Gonçalves, 2005).

The simplest way to estimate VaR and ES is to use the sample quantile estimates based on historic return data – a completely model-free approach. This method is referred to as historical simulation (HS). The HS estimate of VaR_{T+1}^{p} is given by:

$$HS\text{-}VaR_{T+1}^{p} = Q_{1-p}\left(\left\{L_t\right\}_{T-w+1}^{T}\right)$$

where $Q_{1-p}(\bullet)$ is the empirical ($1-p$)-quantile of a moving window of w observations from the distribution of losses L_t. ES is defined accordingly as:

$$HS\text{-}ES_{T+1}^{p} = \frac{1}{\#L_t > HS\text{-}VaR_{t+1}^{p}} \sum_{L_{T+1} > HS\text{-}VaR_{t+1}^{p}} L_t$$

where $\#L_t > HS\text{-}VaR_{t+1}^{p}$ is the number of observations in $\left\{L_t\right\}_{T-w+1}^{T}$ exceeding the VaR calculated through HS.

HS is based on the unconditional distribution of losses, thus, allowing departures from normality in the data to be captured to some extent. Lo (2001), however, stresses that for purposes of active risk management, 'conditional' measures of risk are more relevant than 'unconditional' measures, i.e., VaR and ES calculated through dynamic mean/volatility models vs. HS. Recent articles have also shown that hedge fund returns are often serially correlated (Getmansky et al., 2004) and exhibit dynamic volatility (Giamouridis and Vrontos, 2006). These issues can possibly be addressed by modelling daily losses on a fund through:

$$L_t = \mu_t + \sigma_t \varepsilon_t, \quad t = 1, \ldots, T \tag{1}$$

where location μ_t and scale σ_t are measurable with respect to the information set known at *t-1*, and $\varepsilon_t^{iid} \sim f(\bullet)$, where $f(\bullet)$ is a zero-location, unit-scale probability density with possible additional shape parameters. The conditional mean can be modeled through an ARMA(l,m) specification as follows:

$$\mu_t = \kappa + \sum_{i=1}^{\ell} \varphi_i L_{t-i} + \sum_{j=1}^{m} \theta_j \varepsilon_{t-j} \tag{2}$$

In addition, volatility dynamics can be introduced through a GARCH(r,s) process as follows:

$$\sigma_t^2 = \omega + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2 + \sum_{i=1}^{s} \alpha_i \varepsilon_{t-i}^2 \tag{3}$$

VaR and ES expressions can be then simplified to:

$$VaR_{T+1}^{p} = \mu_{T+1} + \sigma_{T+1} F_{1-p}^{-1} \equiv \mu_{T+1} + \sigma_{T+1} c_{1,p} \tag{4}$$

$$ES_{T+1}^{p} = \mu_{T+1} + \sigma_{T+1} E\left(\varepsilon \middle| \varepsilon > F_{1-p}^{-1}\right) \equiv \mu_{T+1} + \sigma_{T+1} c_{2,p} \tag{5}$$

where F_{1-p}^{-1} denotes the (1-*p*)-th quantile of *f(•)*, the distribution of the standardized losses $\varepsilon_t = \left(L_t - \mu_t\right)/\sigma_t$, and σ_{T+1} is the conditional volatility for time *T*+1, $c_{1,p}$, $c_{2,p}$ are constants depending on *f(•)* and p.

We estimate the parameters of - through a two-stage approach as in McNeil and Frey (2000), Christoffersen and Gonçalves (2005), Kuester et al. (2006) among others. Our approach is summarized as follows:

1. Fit an ARMA(*l,m*)-GARCH(*r,s*) model to the losses L_t data through Gaussian QMLE to obtain $\left[\hat{\kappa}, \left(\hat{\varphi}_1, \ldots, \hat{\varphi}_p\right), \left(\hat{\vartheta}_1, \ldots, \hat{\vartheta}_q\right), \hat{\omega}, \left(\hat{\alpha}_1, \ldots, \hat{\alpha}_r\right), \left(\hat{\beta}_1, \ldots, \hat{\beta}_s\right)\right]$. Compute the conditional mean, $\hat{\mu}_t$, and conditional variance, $\hat{\sigma}_t^2$, sequences through (2) and (3) respectively. Also compute the implied model residual $\hat{\varepsilon}_t = \left(L_t - \hat{\mu}_t\right)/\hat{\sigma}_t$. Finally calculate predictions for $\hat{\mu}_{T+1}$ and $\hat{\sigma}_{T+1}$ through and respectively.

2. Consider the residuals to be a realization of a strict white noise process. Make an assumption for the form of *f(•)*, fit it in the distribution of implied residuals, and compute the estimates of $c_{1,p}$ and $c_{2,p}$, $\hat{c}_{1,p}$ and $\hat{c}_{2,p}$ respectively. Obtain estimates of VaR_{T+1}^{p} and ES_{T+1}^{p} through

$$\widehat{VaR}_{T+1}^{p} = \hat{\mu}_{T+1} + \hat{\sigma}_{T+1} \hat{c}_{1,p} \tag{6}$$

$$\widehat{ES}_{T+1}^{p} = \hat{\mu}_{T+1} + \hat{\sigma}_{T+1} \hat{c}_{2,p} \tag{7}$$

One approach that retains the nonparametric feature of HS and at the same time addresses some of HS's inherent problems, i.e. failure to deal with serial correlation, is the Filtered Historical Simulation (FHS) method. This method is based on estimates of the quantiles from the empirical distribution of the centered residuals. Christoffersen and Gonçalves (2005) list a number of studies which have found the FHS method to perform well, i.e., Hull and White (1998), Barone-Adesi et al. (1999, 2002) and Pritsker (2001). The FHS estimates of VaR_{T+1}^{p} and ES_{T+1}^{p} are computed through (4) and (5) with:

$$\hat{c}_{1,p}^{FHS} = Q_{1-p}\left(\left\{\varepsilon_t - \bar{\varepsilon}\right\}_{t=1}^{T}\right)$$

$$\hat{c}_{2,p}^{FHS} = \frac{1}{\#\left(\varepsilon_t - \bar{\varepsilon}\right) > \hat{c}_{1,p}^{FHS}} \left[\sum_{\left(\varepsilon_t - \bar{\varepsilon}\right) > \hat{c}_{1,p}^{FHS}} \left(\varepsilon_t - \bar{\varepsilon}\right)\right]$$

where $\bar{\varepsilon} = T^{-1} \sum_{t=1}^{T} \varepsilon_t$ to guarantee the zero mean condition.

HS and FHS make no assumption on the probability distribution of the model's residuals. Alternatively, we can assume that *f(•)* is a known parametric density, i.e. Gaussian, with parameters that can be estimated through the sequence of implied residuals. *f(•)* is typically chosen from parametric families which can incorporate features observed in hedge fund return unconditional distributions, i.e. asymmetry and fat tails. Although in a somehow different setting, i.e. unconditional return density modeling, several parametric models have been used in studies concerned with hedge fund risk management. Bali and Gokcan (2004), for example, use the thin-tailed normal distribution, the fat-tailed generalized error distribution, an extreme value approach (a generalized Pareto distribution), and the Cornish-Fisher approximation. Gupta and Liang (2003) and Lhabitant (2001) also use the generalized Pareto distribution. Given the empirical results of the former studies, i.e. that the generalized Pareto distribution is slightly more accurate than the Cornish-Fisher approximation, but the latter might be preferred in practice because of its simplicity (Bali and Gokcan, 2004), and that the generalized Pareto distribution provides a robust and reliable VaR estimation methodology (Gupta and Liang, 2003), we focus on three models: the Gaussian model (G), the Cornish-Fisher model (CF), and the generalized Pareto distribution model (GPD).

Under the assumption that the conditional distribution of ε_t in (1) is Gaussian, the G estimates of VaR_{T+1}^{p} and ES_{T+1}^{p} are computed through and with:

$$\hat{c}_{1,p}^{G} = \Phi_{1-p}^{-1}$$

$$\hat{c}_{2,p}^{G} = \frac{\phi\left(\Phi_{1-p}^{-1}\right)}{p}$$

where Φ and ϕ are the standard cumulative normal, and the standard normal probability, density functions respectively.

CF is based on the principle that the 'true' distribution of ε_t can be approximated with a series expansion of a normal distribution. In particular, the Cornish-Fisher expansion approximates the inverse cumulative density function directly, thus providing the necessary formula for VaR. A Gram-Charlier expansion of a normal density is then used to derive an approximate formula for ES (Giamouridis, 2006). In particular, we can show that CF estimates of VaR_{T+1}^{p} and ES_{T+1}^{p} can be computed through and with:

$$\hat{c}_{1,p}^{CF} = \left\{\Phi_{1-p}^{-1} + \frac{\gamma_1}{6}\left[\left(\Phi_{1-p}^{-1}\right)^2 - 1\right] + \frac{\gamma_2}{24}\left[\left(\Phi_{1-p}^{-1}\right)^3 - 3\Phi_{1-p}^{-1}\right] - \frac{\gamma_1^2}{36}\left[2\left(\Phi_{1-p}^{-1}\right)^3 - 5\Phi_{1-p}^{-1}\right]\right\}$$

$$\hat{c}_{2,p}^{CF} = \frac{\phi\left(\hat{c}_{1,p}^{CF}\right)}{p}\left\{1 + \frac{\gamma_1}{6}\left(\hat{c}_{1,p}^{CF}\right)^3 + \frac{\gamma_2}{24}\left[\left(\hat{c}_{1,p}^{CF}\right)^4 - 2\left(\hat{c}_{1,p}^{CF}\right)^2 - 1\right]\right\}$$

Φ and ϕ are the standard cumulative normal, and the standard normal probability, density functions respectively, and $\gamma_1 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^3$, $\gamma_2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^4 - 3$, where ε_t are the centralized residuals.

GPD is concerned with the tails of the distribution of ε_t. It is aimed to deal with extreme values, thus, making it appropriate for the analysis of hedge fund returns which in most cases involve such values. Let $\varepsilon_{(1)} \geq \varepsilon_{(2)} \geq \ldots \geq \varepsilon_{(n)}$ represent the ordered residuals and suppose that the sufficient – to estimate the parameters of the GPD - number of data in the tail of the distribution is k, $k \ll n$. The GPD with parameters ξ and ϑ is fitted to the data $\left(\varepsilon_{(1)} - \varepsilon_{(k+1)}, \varepsilon_{(2)} - \varepsilon_{(k+1)}, \ldots, \varepsilon_{(k)} - \varepsilon_{(k+1)}\right)$, the excess amounts over the threshold for all residuals exceeding the threshold. It follows (see, e.g. McNeil and Frey, 2000) that VaR_{T+1}^{p} and ES_{T+1}^{p} estimates can be obtained (4) and (5) with:

$$\hat{c}_{1,p}^{GPD} = \hat{\varepsilon}_{(k+1)} + \frac{\hat{\vartheta}}{\hat{\xi}}\left[\left(\frac{p}{k/n}\right)^{-\hat{\xi}} - 1\right]$$

$$\hat{c}_{2,p}^{GPD} = \hat{c}_{1,p}^{GPD}\left[\frac{1}{1-\hat{\xi}} + \frac{\hat{\vartheta} - \hat{\xi}\hat{\varepsilon}_{(k+1)}}{\left(1-\hat{\xi}\right)\hat{c}_{1,p}^{GPD}}\right]$$

3. COMPARING AND TESTING VAR AND ES

This section provides a brief introduction to the procedures used for backtesting VaR and ES predictions computed from the models discussed above. We employ standard methods for VaR backtesting, i.e., unconditional coverage, independence, and conditional coverage (see Christofferesen, 2003 for a detailed overview of these methods). ES is backtested using McNeil and Frey (2000) and Embrechts et al. (2005).

3.1 Unconditional coverage, independence, and conditional coverage testing for VaR predictions

Given a time series of past ex ante VaR forecasts and ex post losses, the 'hit sequence' of VaR violations is defined as

$$I_{t+1} = \begin{cases} 1, & L_{t+1} > VaR_{t+1}^{p} \\ 0, & L_{t+1} < VaR_{t+1}^{p} \end{cases} \qquad (8)$$

Christoffersen (1998) shows that efficiency implies

$$I_{T+1} \overset{iid}{\sim} Bernoulli(p) \tag{9}$$

A test for unconditional coverage investigates if the fraction of violations for a particular model is significantly different from the promised fraction, p. The likelihood-ratio test statistic

$$LR_{uc} = -2\ln\left[(1-p)^{T_0} p^{T_1} \Big/ \left\{(1-T_1/T)^{T_0}(T_1/T)^{T_1}\right\}\right] \overset{asy}{\sim} \chi_1^2 \tag{10}$$

where T, T_0 and T_1 are the total number of observations used for the test, the number of 0s, and the number of 1s, respectively. Monte Carlo simulation is used to obtain P-values in cases of small samples (see Christoffersen, 2003, p.186).

One shortcoming of the test for unconditional coverage is that it cannot reject VaR models which imply clustered violations—mostly attributed to temporal volatility dependence—if the average coverage is correct. Christoffersen (1998) deals with this issue by modeling the 'hit sequence' as a first-order Markov sequence with transition probability matrix

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

where ϖ_{01} denotes the probability of tomorrow being a violation (that is, $I_{t+1} = 1$) conditional on today being a non-violation (that is, $I_t = 0$). ϖ_{11} is defined accordingly. A test for *independence* (H_0: $\varpi_{01} = \varpi_{11}$) is performed through the likelihood-ratio test statistic

$$\mathrm{LR}_{\mathrm{ind}} = -2\ln\left[\mathrm{L}(\hat{\pi}) \Big/ \mathrm{L}\left(\hat{\Pi}_1\right)\right] \overset{\mathrm{asy}}{\sim} \chi_1^2 \tag{11}$$

where Tij, i, j = 0, 1 is the number of observations with a j following an I, $L(\hat{\pi}) = (1-T_1/T)^{T_0}(T_1/T)^{T_1}$ is the likelihood under the alternative hypothesis from the $\mathrm{LR}_{\mathrm{uc}}$ test, and $L\left(\hat{\Pi}_1\right) = (1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}(1-\hat{\pi}_{11})^{T_{10}}\hat{\pi}_{11}^{T_{11}}$ is the likelihood under the alternative hypothesis, with $\hat{\pi}_{01} = T_{01}/(T_{00}+T_{01})$, $\hat{\pi}_{11} = T_{11}/(T_{10}+T_{11})$. When $T_{11} = 0$, $L\left(\hat{\Pi}_1\right) = (1-\hat{\pi}_{01})^{T_{00}}\hat{\pi}_{01}^{T_{01}}$. A Monte Carlo simulation can be used in small samples to simulate appropriate P-values for this test.

A joint test of the above hypotheses—unconditional coverage and independence—can be constructed. This is termed conditional coverage test and is based on the likelihood-ratio test statistic

$$\mathrm{LR}_{\mathrm{cc}} = -2\ln\left[\mathrm{L}(\mathrm{p}) \Big/ \mathrm{L}\left(\hat{\Pi}_1\right)\right] \overset{\mathrm{asy}}{\sim} \chi_2^2 \tag{12}$$

where the P-value of the test can also be computed via Monte Carlo simulation.

3.2 Bootstrap and other procedures for backtesting ES predictions

McNeil and Frey (2000) examine the goodness of ES predictions by means of a bootstrap test. Given a time series of past ex ante ES forecasts and ex post losses exceeding VaR, the series of exceedance residuals can be formulated as

$$\left\{ r_{t+1} = \frac{L_{t+1} - \widehat{ES}_{t+1}^{p}}{\hat{\sigma}_{t+1}} : L_{t+1} > \widehat{VaR}_{t+1}^{p} \right\}$$

Under the null hypothesis that the dynamics of the process and the first moment of the innovation distribution, i.e., $c_{l,p}$, are estimated 'correctly', these residuals should behave like an *iid* sample with mean zero. The mean zero hypothesis can be tested through a bootstrap test that makes no assumption about the underlying distribution of the residuals. The former test examines if an ES prediction model is consistent with its theoretical assumptions. It provides no information, however, on how alternative methodologies rank. Embrechts et al. (2005) demonstrate ES backtesting measures in this direction. The first measure is essentially the conditional average of the losses exceeding ES, in particular

$$V_1^{ES} = \frac{\left\{ \sum_t \left(L_{t+1} - \widehat{ES}_{t+1}^{p}\right) \Big/ L_{t+1} > \widehat{VaR}_{t+1}^{p} \right\}}{\# L_{t+1} > \widehat{VaR}_{t+1}^{p}} \tag{13}$$

suggesting that the more accurate the prediction, the lower, in absolute terms, the value of this measure. A second measure is concerned with the loss incurred in a 1/p-event rather than the behaviour below a certain value, i.e., the VaR. It is defined as

$$V_2^{ES} = \frac{\left\{ \sum_t \left(L_{t+1} - \widehat{ES}_{t+1}^p \right) \Big/ \left(L_{t+1} - \widehat{ES}_{t+1}^p \right) > D_{t+1}^p \right\}}{\#\left(L_{t+1} - \widehat{ES}_{t+1}^p \right) > D_{t+1}^p} \tag{14}$$

where D_{t+1}^p denotes the empirical p-quantile of $\left\{ L_{t+1} - \widehat{ES}_{t+1}^p \right\}$. Again, a good estimation for ES will lead to low absolute values of V_2^{ES}. Finally, the two tests can be combined to formulate

$$V^{ES} = \frac{\left| V_1^{ES} \right| + \left| V_2^{ES} \right|}{2} \tag{15}$$

4. DATA AND EMPIRICAL ANALYSIS

We conduct our analysis with several hedge fund style indices to understand if there are differences in terms of models' accuracy across styles. Rather than disaggregating risk exposures and computing VaR and ES in a multivariate framework we carry out our analysis in a univariate framework—VaR and ES forecasts are derived from the distribution of the fund style's losses as in Bali and Gokcan (2004), Gupta and Liang (2003), Krokhmal, et al. (2002), Lopez de Prado and Peijan (2004). We consider eight HFR investable strategy indices: Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA). Bali and Gokcan (2004) outline the advantages of HFR indices which include minimal survivorship bias, since liquidated funds are sustained in the construction of the indices, increased representation of the universe of hedge funds due to large coverage, and finally, equal contribution of large and small assets under management funds.

The sample covers the period starting March 3, 2003, through to May, 25, 2006, and consists of daily return data. Table 1 presents descriptive statistics of the sample. Given the limited number of available data—796 daily NAV observations—we have very little choice with respect to determining the size of the estimation and out-of-sample periods. We use a reasonably sized moving window of w = 500 observations (Hartz *et al.*, 2006, Kuester et al., 2006, Neftci, 2000). This choice leaves us with 295 observations to evaluate the performance of each model in predicting VaR and ES, the best we can do given the availability of data. Since the length of the out-of-sample period may not be appropriate for the tests described in section 3,[4] we report results from this analysis but refer to them with caution. Our primary discussion is similar with Brooks et al. (2005), Lambadiaris et al. (2003), and Neftci (2000) who address model evaluation more qualitatively, in the spirit of the Base Committee's requirements.

Table 1. Descriptive statistics

	Mean	SD	Median	Q25	Q75	Skew	Kurtosis	Min	Max
EH	0.0255	0.2928	0.0549	-0.1430	0.2219	-.08510	3.7440	-1.1432	0.9406
M	0.0258	0.3527	0.0339	-0.1581	0.2446	0.0463	4.9050	-1.6820	1.1382
RVA	0.0135	0.1817	0.0116	-0.0762	0.1051	0.0318	13.6300	-0.9504	1.6129
ED	0.0369	0.2245	0.0494	-0.0944	0.1716	-0.0810	3.8563	-0.8183	0.7641
CA	0.0002	0.2158	-0.0020	-0.1294	0.1389	0.0502	4.5553	-0.9315	0.8072
DS	0.0336	0.1625	0.0247	-0.0621	0.1166	0.0288	5.5667	-0.7362	0.7672
EMN	-0.0009	0.1694	0.0062	-0.1111	0.1066	-0.0780	3.5677	-0.5764	0.7162
MA	0.0191	0.1749	0.0244	-0.0673	0.1158	-0.0019	5.7901	-0.8686	0.7122

This table presents summary statistics of daily returns for eight HFR indexes from March 3, 2003, through to May, 25, 2006. The summary statistics include mean, standard deviation (SD), median, 25% and 75% quartiles (Q25 and Q75 respectively), skewness (Skew), kurtosis, minimum (Min) and maximum (Max). It is noted that the skewness is measured with the quartile or Bowley skewness coefficient to avoid misinterpretations due to outliers in the data set.

4 - Harmantzis et al. (2005) and Lambadiaris et al. (2003) use similar out-of-sample periods; however, the vast majority of existing studies use over one thousand observations i.e., Hartz, et al. (2006), Kuester et al. (2006), Mc Neil (1999), Mc Neil and Frey (2000), among others.

4.1 Mean and variance models

Our first concern is to determine a tenable ARMA(*l,m*)-GARCH(*r,s*) model for the loss L_t generating process. In McNeil and Frey (2000) a model is considered adequate if no evidence against the *iid*-hypothesis is found. In particular, they examine autocorrelation in the residuals and the absolute residuals through Ljung-Box tests in selected periods. We ran Ljung-Box tests in the entire backtest period. With very few exceptions,[5] we find that the ARMA(*l,m*)-GARCH(*r,s*) specifications in table 2 provide no evidence against the *iid*-hypothesis. In addition, the out-of-sample test discussed in Christoffersen (2003) is performed. The test is based on simple regressions of the squared returns in the forecast period on forecasts from the variance model. It essentially examines if the variance forecast is unbiased. We find that all the retained mean/variance models provide unbiased forecasts for the realized volatility.

Table 2. Conditional mean and variance models

	EH	M	RVA	ED	CA	DS	EMN	MA
Mean								
l	1	-	-	2	2	4	-	4
m	-	-	-	-	3	4	-	4
Variance								
r	1	1	1	1	1	1	1	1
s	1	1	1	1	1	1	1	1

This table presents the number of loss (*l*) and error (*m*) lags in the mean and the variance (*r*) and squared error (*s*) lags used in the variance specification. The strategy indexes considered consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

4.2 Value-at-Risk prediction and backtests

We present the results of 5% VaR calculations in table 3. With 295 out-of-sample data points the expected number of violations, i.e., the observed loss is greater than the predicted VaR, is at most about fifteen (5%x295). In most cases this threshold is exceeded. HS125 and HS250 (that is HS with data windows of 125 and 250 days respectively) appear nearly at all times with the most violations. The average VaR based upon FHS, G, CF, and GPD are almost always higher than that for HS250. This result highlights the excess volatility persistence implied in the GARCH model (see Brooks et al., 2005). HS125 doesn't always seem to be able to address this phenomenon successfully, though. In EH, M, RVA, for example, despite the higher HS125 VaR there are more violations than with the lower average VaR prediction models FHS, G, CF, and GPD. If we take the average size of violations into account[6] we conclude that FHS, G, CF, and GPD outperform HS125 and HS250 given the generally lower, average size of violations, and number of threshold exceedances. Our results do not provide strong evidence in favor of either of FHS, G, CF, and GPD for predicting 5% VaR. Table 4 presents the results of 1% VaR calculations where the expected number of violations is three (1%x295). While all approaches fail to satisfy this criterion, G is the worst—there is no single strategy with less than or exactly three violations—with HS125 being the second worst. While this might have been expected for G, since the average predicted VaR is the lowest compared to other approaches, it is surprising that HS125, the methodology with the highest average predicted VaR, does that badly. With respect to the average loss when the VaR is violated we see that FHS, CF, and GPD perform similarly, with GPD presenting a slightly better performance, followed by CF and FHS. These results are consistent with existing evidence suggesting that GPD and CF perform well in predicting the VaR of hedge fund strategies (Bali and Gokcan, 2004) and with Barone-Adesi et al. (1999, 2002) and Pritsker (2001), who find that FHS compares favorably with the HS approach.

5 - We found evidence against the iid-hypothesis only for the mean model for Convertible Arbitrage (3 violations), and for the mean and the variance models for Merger Arbitrage (5 and 1 violation, respectively).

6 - This measure is in the spirit of Neftci's (2000) mean excess function.

Table 3. Value-at-Risk calculated at 5%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
no	19 (6.44%)	18 (6.10%)	14 (4.75%)	22 (7.46%)	15 (5.08%)	23 (7.80%)	19 (6.44%)	18 (6.10%)
avg.	0.4910	0.5721	0.2021	0.3700	0.4241	0.2386	0.2877	0.3202
avg. size of viol.	0.2641	0.3068	0.0724	0.1403	0.1957	0.0984	0.0768	0.1853
HS250								
no	19 (6.44%)	18 (6.10%)	11 (3.73%)	23 (7.80%)	17 (5.76%)	24 (8.14%)	23 (7.80%)	21 (7.12%)
avg.	0.4702	0.5397	0.2382	0.3404	0.4278	0.2180	0.2773	0.2539
avg. size of viol.	0.2577	0.2997	0.0676	0.1499	0.2287	0.1062	0.0674	0.1854
FHS								
no	14 (4.75%)	13 (4.41%)	8 (2.71%)	19 (6.44%)	20 (6.78%)	21 (7.12%)	20 (6.78%)	21 (7.12%)
avg.	0.5165	0.5782	0.2230	0.3580	0.3429	0.2050	0.2910	0.2655
avg. size of viol.	0.1831	0.3002	0.0988	0.1327	0.1459	0.1136	0.0739	0.1214
G								
no	20 (6.78%)	15 (5.08%)	8 (2.71%)	21 (7.12%)	20 (6.78%)	18 (6.10%)	21 (7.12%)	22 (7.46%)
avg.	0.4539	0.5679	0.2221	0.3311	0.3565	0.2363	0.2688	0.2538
avg. size of viol.	0.1772	0.2659	0.0933	0.1410	0.1252	0.0991	0.0906	0.1275
CF								
no	16 (5.42%)	13 (4.41%)	8 (2.71%)	20 (6.78%)	18 (6.10%)	21 (7.12%)	21 (7.12%)	20 (6.78%)
avg.	0.4803	0.5835	0.2254	0.3320	0.3710	0.2041	0.2712	0.2655
avg. size of viol.	0.1904	0.2899	0.0927	0.1459	0.1250	0.1159	0.0906	0.1268
GPD								
no	14 (4.75%)	14 (4.75%)	8 (2.71%)	19 (6.44%)	20 (6.78%)	21 (7.12%)	19 (6.44%)	22 (7.46%)
avg.	0.5085	0.5683	0.2261	0.3639	0.3449	0.2050	0.2960	0.2612
avg. size of viol.	0.1930	0.2900	0.0947	0.1230	0.1402	0.1150	0.0749	0.1197

This table presents results for the 5% VaR analysis. With 295 observations, the expected coverage for 5% VaR is about 15 observations. The number of violations (no) refers to the times the 5% VaR is exceeded. Figures in brackets compute the proportion relative to the length of the test. The table shows the average 5% VaR (avg.) and the average size of violations, that is bow how much VaR is exceeded on average when it is violated (avg. size of viol.). VaR is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

Table 4. Value-at-Risk calculated at 1%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
no	9 (3.05%)	7 (2.37%)	3 (1.02%)	6 (2.03%)	3 (1.02%)	6 (2.03%)	4 (1.36%)	9 (3.05%)
avg.	0.8293	1.0695	0.3739	0.5547	0.6124	0.4020	0.4050	0.5009
avg. size of viol.	0.1703	0.3132	0.0614	0.1477	0.1353	0.1081	0.0548	0.1068
HS250								
no	8 (2.71%)	3 (1.02%)	1 (0.34%)	8 (2.71%)	2 (0.68%)	6 (2.03%)	6 (2.03%)	6 (2.03%)
avg.	0.7233	0.9682	0.3895	0.5250	0.7418	0.3926	0.3851	0.5194
avg. size of viol.	0.1468	0.5077	0.2129	0.1060	0.2759	0.1024	0.0493	0.2425
FHS								
no	6 (2.03%)	2 (0.68%)	1 (0.34%)	8 (2.71%)	4 (1.36%)	6 (2.03%)	6 (2.03%)	3 (1.02%)
avg.	0.7082	0.9151	0.3685	0.5223	0.5808	0.3497	0.4049	0.4330
avg. size of viol.	0.1366	0.6717	0.1572	0.0409	0.1777	0.0994	0.0585	0.1552
G								
no	8 (2.71%)	6 (2.03%)	4 (1.36%)	9 (3.05%)	4 (1.36%)	4 (1.36%)	6 (2.03%)	6 (2.03%)
avg.	0.6509	0.8066	0.3193	0.4810	0.5030	0.3411	0.3804	0.3679
avg. size of viol.	0.1498	0.2697	0.0533	0.0756	0.2458	0.1301	0.0741	0.1801
CF								
no	6 (2.03%)	2 (0.68%)	1 (0.34%)	8 (2.71%)	4 (1.36%)	5 (1.69%)	6 (2.03%)	3 (1.02%)
avg.	0.7135	0.9889	0.4275	0.5071	0.6132	0.3509	0.4217	0.4563
avg. size of viol.	0.1328	0.5569	0.0263	0.0548	0.1316	0.1116	0.0405	0.1679
GPD								
no	6 (2.03%)	2 (0.68%)	1 (0.34%)	8 (2.71%)	4 (1.36%)	4 (1.36%)	6 (2.03%)	4 (1.36%)
avg.	0.7260	0.9466	0.3942	0.5143	0.5706	0.3503	0.4072	0.4458
avg. size of viol.	0.1278	0.5885	0.1060	0.0487	0.1783	0.1282	0.0559	0.1332

This table presents results for the 5% VaR analysis. With 295 observations, the expected coverage for 1% VaR is about 3 observations. The number of violations (no) refers to the times the 1% VaR is exceeded. Figures in brackets compute the proportion relative to the length of the test. The table shows the average 5% VaR (avg.) and the average size of violations, that is bow how much VaR is exceeded on average when it is violated (avg. size of viol.). VaR is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

Tables 5 and Tables 6 report the finding of the backtesting analysis detailed in section 3.1. For the reasons discussed above these results should be treated with caution. The analysis relies on Monte Carlo simulated P-values. For VaR prediction at 5% FHS, G, CF, and GPD appear to perform equally well. They prove—with insignificant exceptions—to be 'correct' models for predicting 5% VaR for EH, RVA, ED, CA, EMN, and MA in terms of unconditional coverage, independence, and conditional coverage. HS125 is 'correct' for CA and EMN, while HS250 performs really badly. When we examine the 1% VaR the picture is somewhat more diverse. CF and GPD are more often 'correct' than all other models, suggesting that they are more capable of modelling the tails. FHS and HS125 are the second best, followed by HS250 and G.

Table 5. Backtests of Value-at-Risk calculated at 5%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
LR uc	1.1862	0.7065	0.0408	3.2804	0.0044	4.1808	1.1862	0.7065
LRind	12.5440	13.9197	5.3447	5.7734	1.7043	11.6784	0.1848	5.7320
LR cc	13.7302	14.6262	5.3855	9.0537	1.7087	15.8592	1.3710	6.4385
HS250								
LR uc	1.1862	0.7065	1.0963	4.1808	0.3451	5.1756	4.1808	2.4782
LRind	8.4142	13.9197	3.7348	5.0418	10.6208	14.4089	0.6488	6.5752
LR cc	9.6004	14.6262	4.8311	9.2226	10.9659	19.5845	4.8297	9.0534
FHS								
LR uc	0.0408	0.2273	3.8725	1.1862	1.7785	2.4782	1.7785	2.4782
LRind	2.1080	10.8941	1.7050	0.6112	0.4490	6.5752	0.2606	1.5529
LR cc	2.1489	11.1214	5.5774	1.7974	2.2276	9.0534	2.0391	4.0310
G								
LR uc	1.7785	0.0044	3.8725	2.4782	1.7785	0.7065	2.4782	3.2804
LRind	1.9141	8.5010	1.7050	0.3258	0.4490	2.7963	0.3637	1.2402
LR cc	3.6926	8.5055	5.5774	2.8040	2.2276	3.5027	2.8419	4.5206
CF								
LR uc	0.1086	0.2273	3.8725	1.7785	0.7065	2.4782	2.4782	1.7785
LRind	1.3574	10.8941	1.7050	0.4490	0.8147	6.5752	0.3637	1.9141
LR cc	1.4660	11.1214	5.5774	2.2276	1.5211	9.0534	2.8419	3.6926
GPD								
LR uc	0.0408	0.0408	3.8725	1.1862	1.7785	2.4782	1.1862	3.2804
LRind	2.1080	9.6325	1.7050	0.6112	0.4490	6.5752	2.7612	1.2402
LR cc	2.1489	9.6733	5.5774	1.7974	2.2276	9.0534	3.9474	4.5206

This table presents results for the 5% VaR backtest analysis. The null hypothesis tested with LRuc is that the average number of VaR violations is correct. The null hypothesis tested with LRind is that the VaR violations are independent. The null hypothesis tested with LRcc is that the average number of VaR violations is correct and that the VaR violations are independent. P-values are obtained through Monte Carlo simulation (Christoffersen, 2003, p. 186). Figures in bold text indicate failure to reject the null hypothesis at the 10% significance level. VaR is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

Table 6. Backtests of Value-at-Risk calculated at 1%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
LR uc	8.1038	4.0539	0.0009	2.4514	0.0009	2.4514	0.3397	8.1038
LRind	5.2847	7.4190	0.0824	2.7651	0.0824	0.2913	0.1377	0.6307
LR cc	13.3885	11.4729	0.0832	5.2166	0.0832	2.7427	0.4774	8.7345
HS250								
LR uc	5.9500	0.0009	1.7494	5.9500	0.3485	2.4514	2.4514	2.4514
LRind	6.2603	0.0824	0.0136	0.5028	0.0410	0.2913	0.2913	0.2913
LR cc	12.2103	0.0832	1.7630	6.4528	0.3895	2.7427	2.7427	2.7427
FHS								
LR uc	2.4514	0.3485	1.7494	5.9500	0.3397	2.4514	2.4514	0.0009
LRind	0.2913	0.0410	0.0136	0.5028	0.1377	0.2913	0.2913	0.0824
LR cc	2.7427	0.3895	1.7630	6.4528	0.4774	2.7427	2.7427	0.0832
G								
LR uc	5.9500	2.4514	0.3397	8.1038	0.3397	0.3397	2.4514	2.4514
LRind	1.7050	2.7651	4.4559	0.6307	0.1377	0.1377	0.2913	0.2913
LR cc	7.6550	5.2166	4.7956	8.7345	0.4774	0.4774	2.7427	2.7427
CF								
LR uc	2.4514	0.3485	1.7494	5.9500	0.3397	1.1908	2.4514	0.0009
LRind	0.2913	0.0410	0.0136	0.5028	0.1377	0.2073	0.2913	0.0824
LR cc	2.7427	0.3895	1.7630	6.4528	0.4774	1.3981	2.7427	0.0832
GPD								
LR uc	2.4514	0.3485	1.7494	5.9500	0.3397	0.3397	2.4514	0.3397
LRind	0.2913	0.0410	0.0136	0.5028	0.1377	0.1377	0.2913	0.1377
LR cc	2.7427	0.3895	1.7630	6.4528	0.4774	0.4774	2.7427	0.4774

This table presents results for the 1% VaR backtest analysis. The null hypothesis tested with LRuc is that the average number of VaR violations is correct. The null hypothesis tested with LRind is that the VaR violations are independent. The null hypothesis tested with LRcc is that the average number of VaR violations is correct and that the VaR violations are independent. P-values are obtained through Monte Carlo simulation (Christoffersen, 2003, p. 186). Figures in bold text indicate failure to reject the null hypothesis at the 10% significance level. VaR is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

In summary, we have shown that FHS, G, CF, and GPD do almost equally well in predicting 5% VaR and outperform HS125 and HS250 on the basis of lower, average size of violations, and number of threshold violations. The 1% VaR analysis shows that FHS, CF, and GPD perform almost similarly, with GPD presenting with a slightly better performance (CF is second best). HS125 and HS250 despite overestimating VaR present with the highest number and average size of violations. G is the worst approach for predicting 1% VaR. Backtests fail to provide universal evidence in favour of the models examined. Conditional models, i.e. FHS, G, CF, and GPD, prove generally 'correct' for 5% VaR, whereas CF and GPD are in more cases 'correct' for 1% VaR.

4.3 Expected Shortfall prediction and backtests

We present the 5% ES prediction results in table 7. The worst performers in terms of number of violations are HS125 and G. This is not surprising given the relatively few extreme observations in HS125 and the thin tailed density in G. HS250, FHS, CF, and GPD appear with generally similar number of violations. The 1% ES prediction analysis is presented in table 8. The picture is analogous. HS125 and G present highest number of violations, whereas HS250, FHS, CF, and GPD have similar behaviour.

Table 7. Expected Shortfall calculated at 5%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
no	10	8	6	13	9	8	9	11
avg.	0.6787	0.8369	0.2908	0.4881	0.5454	0.3379	0.3577	0.4403
avg. size of viol.	0.2574	0.3782	0.0568	0.1108	0.1221	0.1279	0.0503	0.1513
HS250								
no	10	5	1	10	7	8	9	8
avg.	0.6450	0.8062	0.3425	0.4612	0.6182	0.3268	0.3440	0.4113
avg. size of viol.	0.2092	0.4298	0.2549	0.1458	0.2114	0.1227	0.0610	0.2537
FHS								
no	9	6	3	12	5	13	8	6
avg.	0.6324	0.7937	0.3325	0.4525	0.4896	0.2980	0.3618	0.3727
avg. size of viol.	0.1510	0.2905	0.0661	0.0822	0.2133	0.0718	0.0728	0.1680
G								
no	9	9	5	13	8	11	13	13
avg.	0.5747	0.7142	0.2817	0.4230	0.4463	0.3005	0.3372	0.3238
avg. size of viol.	0.2096	0.2548	0.0769	0.1038	0.1685	0.0756	0.0063	0.1165
CF								
no	8	2	1	12	4	14	8	6
avg.	0.6403	0.8677	0.3613	0.4410	0.5484	0.2683	0.3659	0.4086
avg. size of viol.	0.1571	0.6905	0.1342	0.0923	0.2051	0.0936	0.0702	0.1165
GPD								
no	8	6	2	12	5	13	8	6
avg.	0.6405	0.8033	0.3313	0.4546	0.4916	0.2970	0.3631	0.3757
avg. size of viol.	0.1642	0.2788	0.0973	0.0801	0.2108	0.0719	0.0716	0.1634

This table presents results for the 5% ES analysis. The number of violations (no) refers to the times the 5% ES is exceeded. The table shows also the average 5% ES (avg.) and the average size of violations, that is by how much ES is exceeded on average when it is violated (avg. size of viol.). ES is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

Table 8. Expected Shortfall calculated at 1%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
no	7	4	1	4	2	4	3	6
avg.	0.8890	1.3595	0.4493	0.6670	0.6523	0.4867	0.4350	0.5617
avg. size of viol.	0.1262	0.4652	0.1256	0.1468	0.0984	0.1189	0.0404	0.1002
HS250								
no	3	2	1	4	1	2	2	6
avg.	0.8966	1.3979	0.5279	0.6353	0.8070	0.4684	0.4220	0.6241
avg. size of viol.	0.1502	0.2865	0.1507	0.0842	0.0944	0.1694	0.0521	0.1253
FHS								
no	3	2	0	4	1	1	6	1
avg.	0.7926	1.2050	0.5030	0.5574	0.7700	0.4434	0.4507	0.5736
avg. size of viol.	0.1531	0.3325	N/A	0.0222	0.2947	0.2847	0.0257	0.2128
G								
no	4	2	1	4	4	2	4	5
avg.	0.7488	0.9253	0.3677	0.5555	0.5758	0.3932	0.4360	0.4246
avg. size of viol.	0.1589	0.6135	0.1273	0.0314	0.1726	0.1686	0.0399	0.1462
CF								
no	2	2	1	0	1	1	1	2
avg.	0.8799	1.2189	0.4484	0.6158	0.7477	0.5974	0.5392	0.5550
avg. size of viol.	0.1392	0.2826	0.0526	N/A	0.2804	0.1585	0.0063	0.1334
GPD								
no	3	2	0	4	1	1	6	2
avg.	0.8049	1.1805	0.5030	0.5554	0.7616	0.4541	0.4457	0.5574
avg. size of viol.	0.1433	0.3433	N/A	0.0307	0.3041	0.2715	0.0294	0.1255

This table presents results for the 1% ES analysis. The number of violations (no) refers to the times the 1% ES is exceeded. The table shows also the average 1% ES (avg.) and the average size of violations, that is by how much ES is exceeded on average when it is violated (avg. size of viol.). VaR is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

We backtest the ES models along the lines of section 3.2. table 9 and table 10 tabulate the findings of the analysis according to Embrechts et al. (2005). Given the relatively small number of extreme observations this analysis ranks all models almost equally. McNeil and Frey's (2000) bootstrap analysis sheds more light on the goodness of the ES models. In particular, it classifies HS250 as a 'correct' model for 5% ES in 6 out of 8 strategies, i.e. M, RVA, ED, DS, EMN, and MA. HS125, FHS, and CF are 'correct' models for 5 strategies. GPD is correct for 4 strategies and G is the worst being successful only for 3 strategies. These results are also reported as indicative considering the amount of data needed for bootstrapping techniques. Also, for this reason bootstrap tests for ES at 1% have not been possible to carry out.

Table 9. Backtest of Expected Shortfall calculated at 5%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
V1ES	0.0980	0.0587	0.0118	0.0329	0.0333	0.0000	0.0072	0.0444
V2ES	1.3916	1.5895	0.6024	1.0459	1.1308	0.7343	0.7631	1.0483
V ES	0.7448	0.8241	0.3071	0.5394	0.5821	0.3672	0.3852	0.5463
HS250								
V1ES	0.0722	0.0112	0.0383	0.0243	0.0184	0.0018	0.0022	0.0304
V2ES	1.3371	1.5625	0.5933	1.0147	1.0733	0.7095	0.7402	0.9036
V ES	0.7047	0.7868	0.3158	0.5195	0.5459	0.3556	0.3712	0.4670
FHS								
V1ES	0.0587	0.0768	0.0184	0.0309	0.0163	0.0254	0.0108	0.0118
V2ES	1.4867	1.6180	0.6637	1.0619	1.0413	0.6607	0.7882	1.0214
V ES	0.7727	0.8474	0.3410	0.5464	0.5288	0.3431	0.3995	0.5166
G								
V1ES	0.0547	0.1165	0.0293	0.0469	0.0306	0.0346	0.0272	0.0548
V2ES	1.4092	1.5197	0.6037	1.0245	0.9669	0.6648	0.7571	0.9246
V ES	0.7320	0.8181	0.3165	0.5357	0.4988	0.3497	0.3922	0.4897
CF								
V1ES	0.0266	0.0002	0.0567	0.0343	0.0620	0.0541	0.0016	0.0231
V2ES	1.4996	1.7068	0.7194	1.0487	1.1239	0.6315	0.7900	1.0933
V ES	0.7631	0.8535	0.3880	0.5415	0.5929	0.3428	0.3958	0.5582
GPD								
V1ES	0.0524	0.0481	0.0193	0.0291	0.0186	0.0261	0.0138	0.0010
V2ES	1.5004	1.6300	0.6641	1.0642	1.0437	0.6594	0.7898	1.0219
V ES	0.7764	0.8391	0.3417	0.5467	0.5312	0.3428	0.4018	0.5115

This table presents results for the 5% ES backtest analysis according to Embrechts et al. (2005). Absolute values of the statistics are presented. A good estimation of ES leads to low absolute values of V1ES, V2ES, and VES. ES is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

Table 10. Backtest of Expected Shortfall calculated at 1%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
V1ES	0.0601	0.0984	0.0152	0.0731	0.0612	0.0716	0.0155	0.0542
V2ES	1.6575	2.2204	0.7923	1.2323	1.2193	0.9834	0.8331	1.2600
V ES	0.8588	1.1594	0.4037	0.6527	0.6403	0.5275	0.4243	0.6571
HS250								
V1ES	0.0259	0.0587	0.1507	0.0030	0.0186	0.0004	0.0015	0.1253
V2ES	1.6179	2.1638	0.7936	1.1947	1.2493	0.8728	0.8239	1.1896
V ES	0.8219	1.1113	0.4721	0.5989	0.6340	0.4366	0.4127	0.6574
FHS								
V1ES	0.0583	0.3325	0.0514	0.0042	0.0288	0.0055	0.0257	0.0319
V2ES	1.7266	2.1733	0.9476	1.1919	1.4279	0.8081	0.8839	1.3724
V ES	0.8924	1.2529	0.4995	0.5980	0.7284	0.4068	0.4548	0.7022
G								
V1ES	0.0542	0.1430	0.0037	0.0006	0.1726	0.0774	0.0244	0.1174
V2ES	1.6433	1.7896	0.7291	1.1851	1.1394	0.7564	0.8789	1.0936
V ES	0.8488	0.9663	0.3664	0.5928	0.6560	0.4169	0.4516	0.6055
CF								
V1ES	0.0246	0.2826	0.0526	0.0534	0.0054	0.1305	0.0642	0.0712
V2ES	1.8312	2.1830	0.8476	1.2611	1.3688	0.9726	1.0311	1.3204
V ES	0.9279	1.2328	0.4501	0.6573	0.6871	0.5516	0.5476	0.6958
GPD								
V1ES	0.0477	0.3433	0.0459	0.0071	0.0169	0.0414	0.0294	0.0135
V2ES	1.7587	2.1281	0.9596	1.1853	1.4254	0.8224	0.8767	1.3587
V ES	0.9032	1.2357	0.5027	0.5962	0.7212	0.4319	0.4530	0.6861

This table presents results for the 1% ES backtest analysis according to Embrechts et al. (2005). Absolute values of the statistics are presented. A good estimation of ES leads to low absolute values of V1ES, V2ES, and VES. ES is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

Table 11. Bootstrap test of Expected Shortfall calculated at 5%

	EH	M	RVA	ED	CA	DS	EMN	MA
HS125								
L	0.1186	-0.4712	-0.3654	0.0456	-0.1846	-0.1125	-0.1608	0.6402
U	0.6091	0.8674	0.3299	0.5791	0.6109	0.4892	0.3876	2.0562
HS250								
L	0.0120	-0.0977	-0.4892	-0.2126	0.077	-0.2387	-0.1444	-0.1262
U	0.7384	0.6321	0.6452	0.4215	0.9753	0.5096	0.3302	1.6687
FHS								
L	0.0924	-0.1092	-0.4635	0.2129	-0.0230	-0.1417	-0.1013	0.0442
U	0.8658	1.1327	0.2728	0.7414	1.0635	0.6453	0.3981	1.7426
G								
L	0.1549	0.0980	-0.1296	0.2465	0.3475	-0.0412	-0.0219	0.4090
U	0.7779	1.2244	0.733	0.8464	1.4276	1.0185	0.3632	1.4652
CF								
L	-0.0091	-0.2881	-0.7864	0.2052	0.1542	0.0918	-0.1664	-0.2057
U	0.6978	0.9192	0.1238	0.7376	1.0896	0.7228	0.2661	1.6072
GPD								
L	0.1050	-0.1921	-0.5038	0.1512	0.0544	-0.1276	-0.0757	0.2394
U	0.7727	1.1889	0.3783	0.9212	1.0515	0.6359	0.4195	1.4172

This table presents results for the bootstrap test for 5% ES. The values correspond to the lower (L) and the upper (U) bounds of a confidence interval for the mean of the exceedance residual constructed at the 5% significance level. The null hypothesis is that the mean is zero. The figures correspond to 50,000 bootstraps and 5% level of significance. If the number of simulation increases to 100,000 the results are almost identical. ES is computed through historical simulation based on 125 past observations (HS125), historical simulation based on 250 past observations (HS250), filtered historical simulation (FHS), Gaussian (G), Cornish-Fisher expansion (CF), and generalized Pareto distribution (GPD) parameterizations. The strategy indexes consist of Equity Hedge (HFRXEH, EH for short), Macro (M), Relative Value Arbitrage (RVA), Event-Driven (ED), Convertible Arbitrage (CA), Distressed Securities (DS), Equity Market Neutral (EMN), and Merger Arbitrage (MA).

In summary, we have shown that HS250, FHS, CF, and GPD are superior ES prediction models in the sense that they predict ES which is not violated as often as in HS125 or G. If we combine this result with the findings of the bootstrap analysis, we conclude that HS250, FHS, and CF do generally better than—although marginally—GPD, and HS125, G.

5. CONCLUSION

This article studies the downside risk of hedge fund strategies by means of VaR and ES. We relax the assumptions of constant mean and volatility in hedge fund returns which have been standard in the study of hedge fund downside risk. In addition, we also use flexible distributional assumptions that have been used successfully in the study of traditional investments' downside risk to incorporate asymmetries and fat tails in the loss distribution.

We show that conditional mean/variance models improve our ability to predict VaR. In particular FHS, G, CF, and GPD do almost equally well in predicting 5% VaR and outperform HS125 and HS250 on the basis of lower, average size of violations, and number of threshold violations. In predicting 1% VaR parameterizations allowing asymmetry and fat tails, i.e. CF, and GPD outperform G, HS125, and HS250, and do similarly well with FHS. Our analysis for ES is not that conclusive given the amount of available data. HS250, FHS, CF, and GPD prove superior ES prediction models on the basis of the number of violations. The ES backtests only marginally favour HS250, FHS, and CF. We stress though that the results of our backtest analysis should be viewed as indicative.

REFERENCES

• Agarwal, V., Naik, N. Y. "Risks and portfolio decisions involving hedge funds". *Review of Financial Studies* 2004; 17; 1; 63-98.

• Bali, T. G., Gokcan, S. "Alternative approaches to estimating VAR for Hedge Fund portfolios". In: B. Schachter (Ed.), Intelligent Hedge Fund investing. *Risk Publications: London*; 2004.

• Barone-Adesi, G., Giannopoulos, K., Vosper, L. "VaR without correlations for portfolios of Derivative securities". *Journal of Futures Markets* 1999; 19; 5; 583-602.

• Barone-Adesi, G., Giannopoulos, K., Vosper, L. "Backtesting Derivative portfolios with Filtered Historical Simulation " (FHS). European Financial Management 2002; 8; 31-58.

• Berkowitz, J., O'Brien, J. "How accurate are Value-at-Risk models at commercial Banks?" *Journal of Finance* 2002; 57; 1093–1111.

• "Basel Committee on Banking Supervision and the International Organization of Securities Commissions". Review of issues relating to Highly Leveraged Institutions (HLIs). Available at BIS: http://www.bis.org/publ/bcbs79.pdf; 2001.

• Brooks, C., Clare, A. D., Dalle Molle, J.W., Persand, G. "A comparison of extreme value theory approaches for determining value at risk". *Journal of Empirical Finance* 2005; 12; 339-352.

• Christoffersen, P. "Elements of financial risk management". San Diego, CA: *Academic Press*; 2003.

• Christoffersen, P. "Evaluating interval forecasts". *International Economic Review* 1998; 39; 841–862.

• Christoffersen, P. F., Gonçalves, S. "Estimation risk in financial risk management". *Journal of Risk* 2005; 7; 1-28.

• Embrechts, P., Kaufmann, R., Patie, P. "Strategic Long-Term financial risks: Single risk factors. Computational optimization and applications" 2005; 32; 61-90.

• Getmansky, M., Lo, A. W., Makarov, I. "An econometric model of serial correlation and illiquidity in hedge fund returns". *Journal of Financial Economics* 2004; 74; 529-609.

• Giamouridis, D. "Estimation risk in financial risk management: A correction". *Journal of Risk* 2006; 8; 4.

• Giamouridis, D., Vrontos, I.D. "Hedge fund portfolio construction: A comparison of static and dynamic approaches". *Journal of Banking and Finance* 2006, forthcoming.

• Gupta, A., Liang, B. Do "Hedge Funds have enough capital? A Value-at-Risk approach". Available at SSRN: http://ssrn.com/abstract=424370; 2003.

• Gupta, P., J. Straatman, S.R. Skallsjo, Krol, A. "Risk Management in a Multi-Strategy Framework". Available at SSRN: http://ssrn.com/abstract=835244; 2005.

• Hartz, C., Mittnik, S., Paolella, M.S. "Accurate Value-at-Risk Forecasting Based on the (good old) Normal-GARCH Model". *Working paper* 2006.

• Harmantzis, F., Miao, L., Chien, Y. F. "Empirical study of Value-at-Risk and Expected Shortfall models with heavy tails". Available at SSRN: http://ssrn.com/abstract=788624; 2005.

• Hull, J., White, A. "Incorporating volatility updating into the historical simulation method for VaR". *Journal of Risk* 1998; 1; Fall.

• Investor Risk Committee (IRC). "Hedge Fund Disclosure for Institutional Investors". Available at IAFE: http://www.iafe.org/; 2000, 2001.

• Jeager, L. Managing risk in alternative investment strategies: Successful investing in Hedge Funds and managed Futures. FT-Prentice Hall; 2002.

• Jorion, P. "Risk management lessons from Long-Term Capital Management". *European Financial Management* 2000; 6; 3; 277-300.

• Krokhmal, P., Uryasev, S., Zrazhevsky, G. "Risk management for hedge fund portfolios: A comparative analysis of linear portfolio rebalancing strategies". *Journal of Alternative Investments* 2002; 5; 1; 10-29.

• Kuester, K., Mittik, S., Paolella, M. S. "Value-at-Risk prediction: A comparison of alternative strategies". *Journal of Financial Econometrics* 2006; 4; 1; 53-89.

• Lambadiaris, G., Papadopoulou, L., Skiadopoulos, G., Zoulis, Y. "VaR: history or simulation?" *Risk* 2003; September; 23-26.

• Lhabitant, F.-S. "Assessing market risk for Hedge Funds portfolios". *Journal of Risk Finance* 2001; Spring; 1-17.

• Lo, A. "Risk management for Hedge Funds: Introduction and overview". *Financial Analysts Journal* 2001; November/December; 16-33.

• López de Prado, M. M. and Peijan, A. "Measuring loss potential of Hedge Fund strategies". *Journal of Alternative Investments* 2004; Summer; 7-31.

• "Managed Funds Association (MFA). Sound practices for Hedge Fund managers". Available at MFA: http://www.mfainfo.org/; 2000, 2003, 2005.

• McNeil, A. J. *"Extreme value theory for risk managers. Internal Modelling and CAD II"*; Risk Books; 1999; 93-113.

• McNeil, A. J., Frey, R. "Estimation of tail-related risk measures for heteroscedastic financial time series: An Extreme Value approach". *Journal of Empirical Finance* 2000; 7; 271-300.

• Neftci, S.N. "Value at Risk calculations, extreme events, and tail estimation". *Journal of Derivatives* 2000; Spring; 1-15.

• Pritsker, M. "The hidden dangers of historical simulation". Working paper 2001-27, Federal Reserve Board 2001; Washington DC.

• "The President's Working Group on Financial Markets. Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management". Available at United States Department of Treasury: http://www.ustreas.gov/press/releases/reports/hedgfund.pdf/; 1999.