SYRTO

Systemic Risk Tomography
Signals, Measurements, Transmission channels and Policy Interventions

Hedge Fund Systemic Risk Signals

Roberto Savona

SYRTO WORKING PAPER SERIES
Working paper n. 6 | 2013

This project is funded by the European Union under the 7th Framework Programme (FP7-SSH/2007-2013) Grant Agreement n° 320270

Hedge Fund Systemic Risk Signals[♣]

July 2013

by

Roberto Savona[†]

[♣] The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n 320270 – SYRTO.
[†] Department of Economics and Management, University of Brescia. Address: Dipartimento di Economia e Management, Università degli Studi di Brescia, c/da S. Chiara n° 50, 25122 Brescia (Italy). Tel.: +39 +30 2988557/552. Fax: +39 +30 295814. E-mail: savona@eco.unibs.it.

Hedge Fund Systemic Risk Signals

July 2013

Abstract

In this paper, we realise an early warning system for hedge funds based on specific red flags that help detect the symptoms of impending extreme negative returns and the contagion effect. To do this we use regression tree analysis to identify a series of splitting rules that act as risk signals. The empirical findings presented herein prove that contagion, crowded trades, leverage commonality and liquidity concerns are the leading indicators for predicting worst returns. We not only provide a variable selection among potential predictors, but also assign specific risk thresholds for the selected key indicators at which the vulnerability of hedge funds becomes systemically relevant.

Keywords: Hedge Funds; Dynamic Conditional Correlations; Time-varying beta; Regression Trees; Early Warning System

JEL codes: C11; C58; G01; G10

I. Introduction

In this this paper we realise an early warning system (EWS) for the extreme negative returns of hedge funds based on specific red flags that help detect the symptoms of risky situations that may result in large-scale crises. The key concept of our work conceives excess correlation as the major symptom of contagion. Thus, following Boyson et al. (2010), we inspect hedge fund filtered returns (asset pricing model residuals) in order to reduce the possibility that we attribute to contagion commonality in returns due to exposure to common risk factors. We also rely on Boyson et al. (2010) to define hedge fund extreme negative returns, which are identified as the returns that fall in the bottom 10% of a hedge fund style's monthly returns, and contagion, which is defined as the number of other hedge fund styles that have a worst return in the same month.

To realise the EWS for hedge funds we rely on regression tree (RT) analysis. We develop a risk monitoring system in the spirit of the signal approach (Manasse and Roubini, 2009), which is based on specific splitting threshold values associated with the selected explanatory variables that help detect potential abnormalities in the form of worst hedge fund returns. Our paper is related to Savona (2012), since we use the three-equation system introduced in such an article to estimate the Bayesian time-varying CAPM beta model. However, while Savona (2012) explore how and why the systematic risk exposures of the major hedge fund strategies vary over time, based upon some exogenous variables that hedge fund managers are assumed to use in changing their trading strategies, here the research question is different as well as the methodological innovation. In this paper we focus on time-varying correlations, which are estimated following Alexander (2002), together with other leading indicators for predicting hedge fund worst returns. The main objective is to realize a system of rules of thumb to capture situations of extreme risk, and to do this we implement a novel regression tree algorithm introduced in Vezzoli and Stone (2007) which is well suited to inspect panel data structures. To our knowledge, this is the first study that uses RT to examine systemic risk in hedge funds.

Using data from the CSFB/Tremont indices over the period from January 1994 to September 2008, we find that contagion, crowded trades and leverage commonality are the most important leading indicators of worst hedge fund returns. Furthermore, market and funding liquidity concerns together increase the risk for hedge funds, since risky clusters are signalled when credit spread widens and funds tend to de-leverage. A clinical study of the reasons for the LTCM collapse occurred in 1998 and sub-prime crisis in terms of worst returns suffered by hedge funds suggests that, on one hand, the LTCM collapse was mainly due to extreme commonality in leverage dynamics and higher leverage level, whereas, on the other, the sub-prime crisis was caused by crowded trades together with a substantial drop in leverage commonality due to strong de-leveraging.

The remainder of this paper is organised as follows. Section II discusses the related literature to our work. Section III presents the methodology, while the dataset used in the paper is discussed in Section IV. Section V reports the empirical results and Section VI concludes.

II. Related Literature

Our work is related to several large bodies of work that focus on correlation as the major indicator of systemic risk. Firstly, our paper is complementary to Stein (2009), who emphasises the role of the comovements induced by both the crowded trade and the leverage effects, thus suggesting a way to explore systemic risk that is economically consistent with the new literature on liquidity spirals (Brunnermeier and Pedersen, 2009) and studies of leveraged arbitrageurs (Shleifer and Vishny, 1997; Morris and Shin, 2004). Following this line of reasoning, other papers that explore how hedge funds comove together, especially in times of stress, are complementary to our study. Billio et al. (2012) use correlation to capture the degree of connectivity among financial institutions and its impact in terms of contagion, spillover effects and joint crashes. Boyson et al. (2010) focus on clustering worst returns and, based upon the arguments developed in Bekaert et al. (2005), define hedge fund contagion as the "correlation over and above what one would expect from economic fundamentals". In their view, the clustering of worst returns is conceived as a form of excess correlation, which in turn results in contagion or interdependencies (Forbes and Rigobon, 2002)[*]. Adrian (2007) relies on hedge fund return correlation to proxy for similarities in hedge fund strategies, which is assumed to be a key determinant of the risk of the entire hedge fund industry.

All these studies point to correlation as a measure of connectivity among hedge funds. However, correlation by itself does not necessarily imply systemic risk, since it may reflect common membership and style affiliations, i.e. common risk factor exposure. For this reason, to inspect systemic risk among hedge funds we filter their monthly returns, namely we remove common variation in fund returns using a new asset pricing model that has recently proven to be effective to describe the time-varying risk exposure of hedge funds.

III. Methodology

Three methodological steps are used in this paper: (i) estimating an asset pricing model for hedge funds and then using filtered returns and time-varying beta estimates to (ii) compute the time-varying correlations and (iii) realise the EWS for the hedge fund industry using the RT approach. As mentioned in the introduction and discussed more deeply in the previous section, such a procedure reflects the central importance we attribute to excess correlation, assumed as the major symptom of contagion. Analytically, correlations are computed for (i) filtered returns, in order to measure *crowded trades*; (ii) time-varying betas, to measure the *leverage commonality* connected to systematic risk exposure variations; and (iii) common hedge fund risk factors, thus measuring *risk factor commonality*. We indeed conjecture that contagion could be connected to commonalities in hedge fund strategies (crowded trades), beta dynamics (leverage commonality) and cross-market comovements (risk factor commonality).

[*] Forbes and Rigobon (2002) define significant increases in cross-market comovements as *contagion*, while continued high levels of correlations are defined as *interdependence*.

III.1 Filtered Returns and Time-Varying Betas

Until recently, research on hedge fund returns has focused on regression approaches in which returns are regressed on risk factors that proxy for different trading strategies assuming constant coefficients. However, empirical findings have proven that the risk exposures of hedge funds change significantly over time. As a result, new approaches addressing time-varying parameters have been proposed in order to handle shifts in coefficient estimates (see, for example, Bollen and Whaley, 2010; Patton and Ramodarai, 2012).

In this paper, we use the recent three-equation system implemented in Savona (2012). This is a Bayesian time-varying CAPM beta model conditional upon exogenous variables that hedge fund managers are assumed to use in changing their trading strategies. The reason we refer to this model is twofold. First, it allows the estimation of time-varying systematic risk exposure, which is needed to measure leverage commonality. Second, as documented in Savona (2012), it is a parsimonious model that has been proved to be better than simple multi-factor asset pricing models with constant coefficients, both in- and out-of-sample, thus resulting in better hedge fund filtering returns. Moreover, this also more accurately measures crowded trades.

III.1.1 Three-Equation System

The model used to estimate filtered returns and time-varying betas assumes that hedge fund managers are predominantly focused on a fund-specific style benchmark expressed as a linear combination of the 7+1 risk factors proposed in Fung and Hsieh (2004; 2007a,b) (see Section IV.2), hereafter termed the FH risk factors. The hedge fund-specific style benchmark is simply obtained by regressing hedge fund returns onto these eight explanatory factors and then taking the corresponding expectation. As discussed in Savona (2012), this procedure is followed in order to construct single index style-matched benchmarks to be used in a CAPM context and then explore the time variability of systematic risk exposure (the CAPM beta). To this end, hedge fund managers are assumed to modify their own strategies (the style benchmark) according to some partly observable primitive risk signals (PRSs), which can be viewed as the "impulse variables" that impact on the time variability of systematic risk exposure. In a sense, these PRSs are latent factors that could affect hedge fund returns, but for which the inner mechanism of such a relationship is partly obscured by the complex nature of the dynamic trading rules followed by managers[†].

III.1.2 The Model

The econometric representation of the model used to filter hedge fund returns and estimate time-varying betas is as follows:

$$r_{p,t} = \alpha_p + \beta_{p,t} r_{b,t} + \varepsilon_{p,t} \tag{1}$$

$$\beta_{p,t} = \mu + \phi(\beta_{p,t-1} - \mu) + \Gamma' \mathbf{z}_t + \eta_{p,t} \tag{2}$$

$$r_{b,t} = \Lambda' \mathbf{z}_t + u_{b,t} \tag{3}$$

The first equation describes the excess return behaviour of the hedge fund index, where α_p is a constant, $\beta_{p,t}$ is the time-varying beta, $r_{b,t}$ is the excess benchmark return and $\varepsilon_{p,t}$ is an error term, i.e. the "filtered return". As discussed in the previous section, $r_{b,t}$ is obtained by regressing hedge fund returns onto the 7+1 FH risk factors and then taking the corresponding expectation expressed in terms of excess returns over the risk-free rate.

The second equation is the single beta relative to the regression-based style benchmark with ϕ to denote the persistence beta parameter, μ the unconditional mean-reverting beta term, Γ' the transposed vector of sensitivities towards $\mathbf{z}_t$, which is the vector of the PRSs, and $\eta_{p,t}$ the stochastic component. The third equation is the fund-specific style benchmark excess return, which is modelled using the same set of covariates used to describe the beta evolution. Λ' is the transposed vector of sensitivities towards $\mathbf{z}_t$ and $u_{b,t}$ is the unexpected benchmark return.

[†] The proxies for these PRSs are discussed in Section IV.3.

In order to derive scale-independent coefficient estimates for PRS sensitivities, we standardised each of the four instruments.

The three-equation system was estimated following a Bayesian approach within a state space technology, in which posterior distributions were simulated by running the Gibbs sampling data augmentation procedure, and time-varying CAPM betas were estimated with the Kalman filter[‡].

III.2 Time-Varying Correlations

To scrutinise the time evolution of crowdedness in trading strategies, together with leverage and risk factor commonality, we rely on dynamic conditional correlations (DCCs) following Alexander (2002). The method is based on GARCH volatilities of the first few principal components of a specific system of factors. We then use the corresponding factor weights to generate correlations of the original system.

We rely on DCCs because a number of studies have concluded that correlations among international market returns are not constant over time (see, for example, Longin and Solnik, 1995; Berben and Jansen, 2005). These changes in correlation are expected to impact on hedge fund returns, as recently discussed in Buraschi et al. (2012). Indeed, hedge funds move market betas and hedge ratios in order to meet their absolute return objectives.

These reasons suggest relying on time-varying conditional correlations using GARCH-based class models, which also allows us to capture complex dynamics in volatility patterns, such as volatility clustering (Lahrecha and Sylwester, 2011) and also other stylized facts in the literature such as persistence and autocorrelations (Christodoulakis and Satchell, 2002).

III.2.1 Correlation Model

Let $\boldsymbol{Y}$ denote a $T \times k$ matrix of data (in our case, filtered returns, time-varying betas and risk factor returns) and let us assume we can extract uncorrelated g principal components with $g < k$. Let a be

[‡] See Savona (2012) for further technical details on this point.

the $g \times 1$ vector of normalised factor weights for hedge fund indices and let d_t be the $g \times 1$ vector of conditional variances in t estimated using the univariate GARCH(1,1) for the first g principal components. To compute the correlations between i and j with $i \neq j$ at time t we run the following equation:

$$\rho_{i,j,t} = \frac{a_i' d_t a_j}{\left[\left(a_i' d_t a_i\right)^{0.5}\right]\left[\left(a_j' d_t a_j\right)^{0.5}\right]}. \tag{4}$$

Equation (4) was computed for:

- Filtered returns of hedge fund indices in order to measure crowded trades;
- Time-varying betas of hedge fund indices to measure leverage commonality; and
- Common hedge fund risk factors (7+1 FH risk factors), thus measuring risk factor commonality.

After having estimated all pairwise correlations for filtered returns, time-varying betas and common risk factors, we next compute the cross-sectional median of the estimated dynamic correlations. This is our measure to describe the tendency of each index to comove with the hedge fund industry as a whole in terms of crowded trades and leverage commonality as well as to describe market comovement dynamics. Note also that the cross-sectional median allows us to control for non-normal cross-distributions (heavy-tailed), thus giving a more efficient measure to what we mean by the "general tendency" of comovement intensity.

III.3 An EWS for the Hedge Fund Industry

The risk monitoring system we propose for hedge funds is obtained through RT analysis. This approach is a statistical technique introduced in Breiman et al. (1984) with the objective of recursively partitioning the predictor space using a series of subsequent nodes that collapse into distinct partitions in which the distribution of the dependent variable, Y, minimises the prediction error within each region. In a sense, it can be viewed as a collection of piecewise linear functions defined by disjoint regions wherein observations are grouped.

RT is well suited to uncovering forms of non-linearity and provides a general non-parametric way of identifying multiple data regimes from a set of predictor variables. This approach has increasingly been applied in the context of financial crisis studies (see, for example, Manasse and Roubini, 2009; Savona and Vezzoli, 2012) to study the complex and non-linear nature of financial crises as well as to realise EWSs with the aim of signalling impending crises when preselected leading economic indicators exceed specific thresholds. In our study we implement a particular regression tree algorithm, the CRAGGING (CRoss-validation AGGregatING), introduced in Vezzoli and Stone (2007) and recently used in Savona and Vezzoli (2012) which was proposed to remove some of the problems of the regression tree analysis when dealing with panel data and other types of structured data. The technical description of the CRAGGING, which is outside the scope of this paper, is presented in Annex B of the Appendix, while in the following we discuss about general issues involving the regression tree analysis, which are instead more relevant to better understand analysis and results of our work.

III.3.1 Technical Issues

Let $\mathbf{X} = [(X_1, \ldots, X_r)]$ be a collection of r vectors of predictors, both quantitative and qualitative. Let T denote a tree with $m = 1, \ldots, M$ terminal nodes, i.e. the disjoint regions $\widetilde{T}_m$, and by $\Theta = \theta_1, \ldots, \theta_M$ the parameter that associates each m-th θ value with the corresponding node. A generic dependent variable Y conditional on Θ assumes the following distribution

$$f(y_i|\Theta) = \sum_{m=1}^{M} \theta_m I\left(\mathbf{X} \in \widetilde{T}_m\right) \tag{5}$$

where θ_m represents a specific $\widetilde{T}_m$ region and I denotes the indicator function that takes the value of 1 if $\mathbf{X} \in \widetilde{T}_m$. This signifies that predictions are computed by the average of the Y values within the terminal nodes, i.e.

$$\hat{y}_i = \hat{\theta}_m \Rightarrow N_m^{-1} \sum_{\mathbf{x}_i \in \widetilde{T}_m} y_i \tag{6}$$

with $i = 1, \dots, N$ the total number of observations and N_m the number within the *m*-th region.

Computationally, the general problem for finding an optimal tree is solved by minimising the following loss function[§]

$$\underset{\Xi=\{T,\Theta\}}{\arg\min} L = \left[Y - f(Y|\Theta)\right]^2 . \tag{7}$$

This entails selecting the optimal number of regions and corresponding splitting values.

Let s^* be the best split value and $R(m) = N_m^{-1} \sum_{\mathbf{x}_i \in \widetilde{T}_m} \left(y_i - \hat{\theta}_m\right)^2$ be the measure of the variability within each node. Thus, the fitting criterion is given by

$$\Delta R\left(s^*, m\right) = \max_{s^*} \Delta R(s, m) \tag{8}$$

with

$$\Delta R(s, m) = R(m) - \left[R(m_1) + R(m_2)\right]. \tag{9}$$

The procedure is run for each predictor. Then, all the best splits on each variable are ranked according to the reduction in impurity achieved by each split. The selected variables and corresponding split points are those that most reduce the loss function in each partition.

Another interesting feature of RTs is that they aim to improve out-of-sample predictability. The estimation process is based on cross-validation, through which the data are partitioned into subsets such that the analysis is initially performed on a single subset (the training sets), while the other subset(s) are retained for subsequent use in confirming and validating the initial analysis (the validation or testing sets).

The properties of the RT approach seem to be a useful way to inspect hedge fund tail events. Indeed, (i) RTs allow for non-linear relationships and predictors can be quantitative or qualitative, thus detecting and revealing interactions in the dataset; (ii) the number of nodes as well as the corresponding splitting threshold values are the output of an optimisation procedure that delivers

[§] In solving such a minimisation process a common procedure is to grow the tree and then control for the overfitting problem by pruning the largest tree according to a cost-complexity function that modulates the trade-off between the size of the tree and its goodness of fit to the data.

the best aggregation of data within homogeneous clusters; and (iii) the procedure is essentially a forecasting model that trades off between fitting and forecasting ability.

RT models are therefore particularly suitable for developing pragmatic EWSs, as they are based on a collection of binary rules of thumb such as "$x_{ji} \leq s_j$" and "$x_{ji} > s_j$" for each j predictor. Indeed, what you get through RT is a risk stratification conceived with the aim of detecting potential extreme risk situations.

In our study, the response variable is the Worst Return (*WR*), which is defined according to Boyson et al. (2010), namely a dummy variable assuming 0 for no *WR* and 1 if we observe an extreme negative return falling within the 10% of the left side of the return distribution of a given index.

Following this approach, we thus realise a sort of rating system based on a series of "red flags" for crowded trades, leverage and risk factor commonality together with other potential predictors. These red flags are assembled together in order to predict an impending worst return, giving the corresponding probability (the average number of worst returns over the total cases classified within each terminal node),

$$\Pr(WR_i) \approx \hat{y}_i = \hat{\theta}_m \Rightarrow N_m^{-1} \sum_{\mathbf{x}_i \in \tilde{T}_m} WR_i \ . \tag{10}$$

IV. Data

IV.1 Hedge Fund Index Returns

The data used for hedge fund styles are the monthly returns of the CSFB/Tremont indices over the period January 1994–September 2008 (summary statistics are in Annex A of the Appendix). These are 10 asset-weighted indices of funds with a minimum of \$10 million of AUM, a minimum one-year track record and current audited financial statements, including Convertible Arbitrage, Dedicated Short Bias, Emerging Markets, Equity Market Neutral, Event Driven, Fixed Income Arbitrage, Global Macro, Long/Short Equity, Managed Futures and Multi-Strategy. To avoid redundancies we do not consider the aggregate index computed from the CSFB/Tremont database nor the three sub-indices of Event Driven Index (Event Driven–Distressed, Event Driven–Multi-Strategy, Event Driven–Risk Arbitrage).

The time period used to inspect worst returns was split into two intervals, the first from January 1998 to December 2006 and the second from January 2007 to September 2008. The first sub-sample was used to estimate the DCCs and our EWS, thus using the second sub-sample as an out-of-sample test set.

Following Savona (2012), the asset pricing model used to estimate filtered returns and time-varying betas was estimated over the same estimation period (January 1998 to December 2006), with the time interval January 1994 to December 1997 as a "pre-sample" for priors' estimation according to the Bayesian approach outlined above. To better inspect the impact of the sub-prime crisis, filtered returns and betas over the period January 2007–September 2008 were computed using the model estimated in-sample. In so doing, we actually reduced the potential bias induced by the model parameters that otherwise would have been obtained if they had been estimated by extending the period to September 2008.

IV.2 Systematic Risk Factors

The risk factors used to estimate the fund-specific style benchmark through a constant multi-beta model are the FH 7+1 risk factors. These authors suggest using: (i) three primitive trend-following strategies proxied as pairs of standard straddles and constructed from exchange-traded put and call options as described in Fung and Hsieh (2001); (ii) two equity-oriented risk factors; (iii) two bond-oriented risk factors; and (iv) one emerging market risk factor. The following indices were used in the empirical analysis: (1) Bond Trend-Following Factor; (2) Currency Trend-Following Factor; (3) Commodity Trend-Following Factor; (4) Standard & Poor's 500 index monthly total return; (5) Size Spread, proxied by Wilshire Small Cap 1750 minus Wilshire Large Cap 750 monthly returns; (6) 10-year Treasury constant maturity yield month end-to-month end change; (7) Credit Spread, proxied by the month end-to-month end change in the Moody's Baa yield less the 10-year treasury constant maturity yield; and (8) MSCI Emerging Market Index (summary statistics are in Annex A of the Appendix).

IV.3 PRSs

As outlined in Section III, PRSs are observable variables that managers are assumed to use in changing their trading strategies and that enter into the beta and benchmark equations. To select appropriate variables for PRSs, Savona (2012) refers to empirical and theoretical papers involved in the issue of risks in hedge funds. The following are the selected variables: (1) CBOE Volatility Index (VIX); (2) Month end-to-month end change in the three-month T-bill; (3) Term Spread, computed as the monthly difference between the yield on 10-year Treasuries and three-month Treasuries; and (4) Innovations in S&P's 500 monthly standard deviation (Inn) as a proxy for liquidity shocks and estimated by the equation $v_t - v_{t-1} = c_v(v_{t-1} - v_f) + s_t$, with v_t and v_{t-1} the market volatility at time t and $t-1$, respectively, and c_v the persistence volatility parameter that shrinks the volatility process towards the long-run fundamental volatility v_f, which assumed to be constant; s_t is the error term used to proxy for liquidity shocks. As discussed in Savona (2012), the rationale of this variable comes from Brunnermeier and Pedersen (2009), who provide a model that relates asset market liquidity to traders' funding liquidity. Among the implications of the Brunnermeier–Pedersen model, these authors point out that market liquidity declines as fundamental volatility increases (summary statistics are in Annex A of the Appendix).

V. Analysis and Results

The next sections describe and comment on the main results obtained in our analysis, which are structured in order to answer the following questions: (i) Are crowded trades together with leverage and common risk factor commonalities, and other factors such as liquidity shocks, responsible for the increase in systemic risk in the hedge fund industry? (ii) Which values for excess correlations and other potential predictors for extreme negative returns should be considered as risk alarm thresholds? (iii) Are we able to put together all such predictors in order to get an EWS to fit and predict past and future hedge fund extreme events that could propagate within the industry? In the next sections we will give an answer to each question.

V.1. An EWS for Extreme Negative Returns

Through the EWS, our aim is to both explain and predict when and why hedge fund styles could experience an extreme negative return. After having estimated and commented on the DCCs as proxies for crowded trades, leverage commonality and risk factor commonality by considering other potential predictors, the next objective is to realise a collection of thresholds to best stratify the potential risk for single hedge fund styles. To do this, we use 30 potential predictors pertaining to the following categories:

- DCCs for filtered returns, time-varying betas and risk factors. We both considered the level of the three DCCs as well as their monthly differences in order to capture sudden changes possibly induced by systemic risk impacts.
- Time-varying betas for each index and their monthly differences;
- Risk factor volatility (*V*) proxied by the cross-sectional weighted average conditional standard deviation of the 7+1 FH risk factors[**]. Level and monthly changes have been considered;
- The 7+1 FH risk factors and four PRSs;
- Hedge fund illiquidity measure introduced in Getmansky et al. (2004)[††] (Lo_ill). Level and monthly changes have been considered;

[**] This variable was computed as the cross-sectional weighted average conditional standard deviation of the 7+1 FH risk factors, using the time-varying standard deviations estimated before through the univariate GARCH (1,1) and the portions of the variance per component as weights. Mathematically, $V_t = \sum_i w_{i,t}\sigma_{i,t}$,where $w_{i,t}$ is the portion of variance for the *i*-th component at time *t*, i.e., the eigenvalue of the factor *i* over the total eigenvalues of the components extracted (for the 7+1 FH risk factors, we extracted four components and the total eigenvalues were then computed as the sum of the eigenvalues of these components before expressing in relative terms the single eigenvalues); $\sigma_{i,t}$ is the conditional time-varying standard deviation for the factor *i* at time *t*.

- AUM by single hedge fund style to proxy for the dimension of funds;
- The Pastor and Stambaugh (2003) measures used to proxy for US market liquidity, namely: (a) the levels of aggregate liquidity, which is a non-traded liquidity factor associated with temporary price fluctuations induced by order flow (PS1); (b) the innovations in the levels of aggregate liquidity factor (PS2); and (c) the traded liquidity factor computed as the value-weighted return on the 10-1 portfolio from the 10 sized portfolios sorted on historical liquidity betas (PS3).
- Contagion (*C*), measured as the number of other hedge fund styles experiencing a worst return in the same month (since we have 10 indices, the values range from 0 to 9).

All predictors were lagged by one month in order to estimate the expected probability of *WR* at time t given the values of predictors observed in $t-1$. For contagion, we used both C measured at time t as well as C at $t-1$ because contagion essentially indicates the switching regimes in the data, and RTs are well suited to endogenously detecting such changes induced by the contagion effect. As a result, it is necessary to consider the value for contagion at the same time as the dependent variable.

We pooled the data from the 10 hedge fund indices and predictors based on the two time intervals January 1998–December 2006 and January 2007–September 2008. As discussed previously, we used the period January 2007–September 2008 to perform the out-of-sample analysis based on models estimated in the period January 1998–December 2006. Moreover, to focus more closely on the two major systemic events during the time period inspected, we also estimated the models for the sub-periods 1998–1999 and 2007–September 2008.

[††] The variable is obtained by the cross-sectional weighted average first-order autocorrelations using a rolling window of 36 past monthly returns and the relative AUM as weights: $\rho_t = \sum_i w_{i,t}\rho(1)_{i,t}$, in which $\rho(1)_{i,t}$ is the average first-order autocorrelation for index *i* at time *t*.

V.2.1 Hedge Fund Risk Stratification

The RT run over the period January 1998–December 2006 stratified the risk of having an extreme negative return in 10 clusters as depicted in Figure 1. The procedure selected six out of 30 predictors: 1) Contagion; 2) DCC Filter; 3) Credit Spread; 4) Change in Beta; 5) DCC Beta; and 6) AUM. This result seems to empirically confirm and extend the arguments developed in Stein (2009), since crowded trades (DCC Filter) and liquidity concerns (Credit Spread changes)[‡‡] together with leverage dynamics (commonality, DCC Beta and change in level, $d\beta$) and the hedge fund style dimension (AUM) contribute to explaining the worst returns, especially in times of contagion.

An in-depth exploration of the partitions realised through our analysis allowed us to identify the following risk levels:

- **Extreme Risk**, signalled when the contagion effect ($C_t \geq 3$) is associated with high leverage commonality ($\text{DCC Beta} > 0.6933$) and the style dimension is alternatively high ($\text{AUM} > 0.093$) or low ($\text{AUM} \leq 0.036$). The risk is slightly higher for smaller funds for which we estimate $\Pr(WR) = 0.8557$ (node #8) against $\Pr(WR) = 0.7159$ (node #10) for larger ones.
- **High Risk**, namely when substantial leverage commonality ($\text{DCC Beta} > 0.6933$) is connected alternatively with (a) systematic risk reduction (de-leveraging) ($d\beta \leq -0.26231$) or (b) median dimension-based funds ($0.036 < \text{AUM} \leq 0.093$) during times of contagion ($C_t \geq 3$). The probability estimates for (a) and (b) are $\Pr(WR) = 0.4443$ (node #6) and $\Pr(WR) = 0.3315$ (node #9), respectively.
- **Medium Risk**, namely when crowded trades are significantly negative, i.e. when proprietary trading strategies are, in some senses, opposite to other competitors ($\text{DCC Filter} \leq -0.263$),

‡‡ Credit Spread can be viewed also as a proxy for funding liquidity risk faced by hedge funds. Patton and Ramodarai (2012), for example, use this variable to capture variation in the availability of credit on account of changes in the probability of default.

which is the case for some Dedicated Short Bias funds[§§]. For the funds we define as Strong Short Bias, we estimate $\Pr(WR)=0.2684$ (node #1). Moreover, a similar risk level is signalled for all the other funds, i.e. those having ($\text{DCC Filter} > -0.263$), whenever credit spread widens ($\text{CS} > 6.5\text{bp}$) together with substantial de-leveraging ($d\beta \le -0.2086$), which seems to delineate a situation in which market illiquidity (implied in widened credit spreads) forces hedge funds to reduce their leverage levels. In this case, we have $\Pr(WR)=0.2271$ (node #3).

- **Moderate Risk**, for funds displaying low commonality in leverage dynamics ($\text{DCC Beta} \le 0.6933$) and showing substantial leverage commonality ($\text{DCC Beta} > 0.6933$) with no extreme de-leveraging ($d\beta > -0.26231$). The probability estimates are, in order, $\Pr(WR)=0.1561$ (node #5) and $\Pr(WR)=0.0831$ (node #7).
- **Low Risk**, when Credit Spread does not widen significantly ($\text{CS} \le 6.5\text{bp}$) and funds are not in the Strong Short Bias style ($\text{DCC Filter} > -0.263$). In this case, we have the lowest probability of suffering from a *WR* with $\Pr(WR)=0.0388$ (node #2). Alternatively, the same risk level occurs when positive Credit Spread changes ($\text{CS} > 6.5\text{bp}$); again the style is not Strong Short Bias ($\text{DCC Filter} > -0.263$) and funds tend to increase their systematic risk exposure ($d\beta > -0.2086$). Here, the probability estimate is $\Pr(WR)=0.06757$ (node #4). The fact that Credit Spread is connected to changes in beta seems to suggest that the predictor could indicate liquidity concerns when linked to fund de-leveraging. Indeed, the threshold for $d\beta$ discriminates between moderate risk, when $d\beta \le -0.2086$, and low risk, for $d\beta > -0.2086$.

The main conclusion of this analysis is that contagion and leverage commonality are leading indicators in signalling extreme risk situations. Having three or more fund styles experiencing an

[§§] All funds clustered within this node were Dedicated Short Bias.

extreme negative return and following strategies that imply significant communality in beta dynamics, i.e. greater than ≅ 0.7, leads to a higher probability of having a worst return. Interestingly, the time series that are located within the two higher risk clusters include the periods August–October 1998, September 2001 and April–May 2005, namely the LTCM collapse, terrorist attacks of 09/11, and Ford and GM downgrade, respectively. Liquidity concerns seem to move in tandem with changes in leverage, since they lead to risky clusters when credit spread widens and funds tend to de-leverage.

V.3 Robustness Checks

In order to assess the robustness of our EWS, we check the model performance for the in-sample and out-of-sample and verified the consistency of the results of our model with those from Logit specifications.

V.3.1 In-Sample and Out-Of-Sample Model Accuracy

The model accuracy was evaluated using common scoring- and signal-based diagnostic tests. The first is the Brier Score (*BS*), which is the average squared deviation between predicted probabilities and actual outcomes, assigning a lower score for higher accuracy,

$$BS = N^{-1} \cdot \sum_i 2(\hat{y}_i - y_i)^2 \qquad BS \in [0, 2]. \tag{11}$$

where N is the number of observations and $\hat{y}_i$ and y_i are the estimated and actual response variables for the i-th observation.

Secondly, we rely on signal-based diagnostic tests using the ROC curve. These include: (1) the Youden Index, which is a summary measure of the model accuracy considering both type-I and type-II errors commonly used to find the optimal cut-off point in classification studies (predicting *WR*, 1, and no-*WR*, 0). The measure is computed as $[(1-\alpha)+(1-\beta)-1]$ where α and β are type-I and type-II errors; (2) the optimal cut-off point, corresponding to the value of the probability estimate $\hat{y}$, to be used to classify predictions into a 1-0 variable, which maximises the Youden Index; (3) Sensitivity, which is the ratio of *WR* correctly classified over the actual *WR*, $(1-\alpha)$; (4) Specificity, which is the ratio of no-*WR* correctly classified over the actual no-*WR*, $(1-\beta)$; and (5) the Area Under the ROC Curve (AUC), which is a measure of the model classification ability ranging from 0 (random model with no classification ability) to 1 (perfect model) and the AUC, which is in turn a function mapping sensitivity onto type-II errors for each possible threshold; we then visualise the trade-off between type-I and type-II errors.

The results reported in Table 1 show that the overall performance of the EWS as measured by the AUC as well as sensitivity and specificity computed using the optimal cut-off points through the Youden Index denote significant changes for the in-sample and out-of-sample. Indeed, looking at type-I errors, we note that the model predicts 59 out 90 *WR*s in-sample (i.e., a sensitivity of 0.6556), while the out-of-sample model predicts 33 out 36 *WR*s with a sensitivity of 0.9167. By contrast, specificity is 0.7340 for the in-sample and 0.6149 for the out-of-sample. In other terms, the EWS modulates classification errors and shows a higher ability in predicting *WR*s (true positive) in the out-of-sample to the detriment of specificity, since false alarms increase from the first to the second time period. This is the main reason why the Brier Score changes from 0.1412 (in-sample) to 0.3004 (out-of-sample), thus indicating a deterioration in model accuracy assessed in the holdout period.

To summarise, although the performance of the EWS in-sample is enough, with substantial missed *WR*s, out-of-sample the model is extremely good at predicting whether hedge funds will experience an extreme negative return. Indeed, the AUC in- and out-of-sample is 0.7294 and 0.8554, respectively.

Additional results on the robustness of the regression tree algorithm used in our study compared with alternative approaches are reported and commented in Annex C of the Appendix, in which we prove that out-of-sample our EWS is the best performer compared with the original regression tree approach introduced in Breiman et al. (1984) and the more recent procedure introduced in Hothorn et al. (2006), to cope with the overfitting and variable selection problems induced by a recursive fitting procedure.

V.3.2 Logit-Based Consistency Comparison

To reconcile and confront the results of our non-parametric model with those from traditional parametric approaches, we run a Logit model using the covariates selected by the RT and including the interaction terms relative to the primary splitters. This approach is similar in spirit to that of Dattagupta and Cashin (2012), although the way in which they reconcile the RT with the Logit is different from our approach. Indeed, Dattagupta and Cashin (2012) reconcile the Logit and RT results by including dummy variables for the primary splitters at the thresholds identified by the baseline RT model. Instead, we incorporate the non-linear effects of the primary splitters by including the interactions among the corresponding covariates. Indeed, by observing Figure 1 we note major interactions with: (i) commonality in filtered returns (DCC Filter), credit spread (CS) and change in betas ($d\beta$), when contagion is less than 1; and (ii) leverage commonality (DCC Beta) and change in betas ($d\beta$), when contagion is greater than 1.

The following reconciling Logit model was thus used:

$$
\begin{aligned}
\mathrm{Logit}\left[P\left(y_{it}=1\right)\right] = {} & b_0 + b_1 \cdot C_t + b_2 \cdot \mathrm{DCC\,Beta}_{t\text{-}1} + b_3 \cdot \mathrm{DCC\,Filter}_{t\text{-}1} + b_4 \cdot d\beta_{t\text{-}1} + b_5 \cdot \mathrm{AUM}_{t\text{-}1} \\
& + b_6 \cdot \mathrm{CS}_{t\text{-}1} + b_7 \cdot d_1 \cdot \left(C_t \cdot \mathrm{DCC\,Filter}_{t\text{-}1} \cdot \mathrm{CS}_{t\text{-}1}\right) + b_8 \cdot d_2 \cdot \left(C_t \cdot \mathrm{DCC\,Beta}_{t\text{-}1}\right) \\
& + b_9 \cdot d_2 \cdot \left(C_t \cdot \mathrm{DCC\,Beta}_{t\text{-}1} \cdot d\beta_{t\text{-}1}\right)
\end{aligned}
$$

with (12)

$$
d_1 \begin{cases} 1 & \text{for } C_t \leq 1 \\ 0 & \text{otherwise} \end{cases}
$$

$$
d_2 \begin{cases} 1 & \text{for } C_t > 1 \\ 0 & \text{otherwise} \end{cases}
$$

where d_1 and d_2 are the dummy variables used to take into account the major interactions of the RT when $C_t \leq 1$ (nodes 1 to 4) and $C_t > 1$ (nodes 5 to 10).

The estimation results are in Table 2, in which we report the Logit model with interaction terms (Logit 1) and without interactions, i.e. only considering the RT covariates (Logit 2). Using the base Logit specification (Logit 2), a significant increase in *WR* probability is associated with: (i) an increase in contagion; (ii) high hedge fund dimension (AUM); and (iii) low credit spread. These results only partially agree with those of RT (having only for contagion the expected direction of the coefficient and only three out six significant explanatory variables). In fact, the high hedge fund dimension and low credit spread associated with an increase in *WR* probability do not seem to be in line with our EWS.

When instead we include the interaction terms, the Logit specification (Logit 1) denotes five out nine significant explanatory variables. However, in this case we also have contrasting signs for some estimated coefficients. Indeed, an increase in *WR* probability is associated with: (i) high contagion, which is coherent with our expectation; (ii) low leverage commonality, which contrasts with RT; and (iii) high fund dimension and low credit spreads, which are somewhat difficult to compare with the RT results since they have non-linear relationships with other covariates. Furthermore, the reduction in systematic risk exposure ($d\beta$), which is in line with RT, seems to be near to significant.

The interaction terms are statistically significant, except for $d_1 \cdot \left(C_t \cdot \mathrm{DCC\,Filter}_{t\text{-}1} \cdot \mathrm{CS}_{t\text{-}1}\right)$, thus supporting the RT results relative to the non-linear relationships among contagion, leverage commonality and change in betas: $d_2 \cdot \left(C_t \cdot \mathrm{DCC\,Beta}_{t\text{-}1}\right)$ and $d_2 \cdot \left(C_t \cdot \mathrm{DCC\,Beta}_{t\text{-}1} \cdot d\beta_{t\text{-}1}\right)$ show both positive coefficients, thereby confirming the leading role of contagion, leverage commonality and systematic risk exposure splitters (nodes 5 to 10).

The above results clearly show that only by considering interactions among predictors can the Logit specification reconcile with RT, although only major non-linearity can be captured. On this point, the non-parametric recursive partition provides a more detailed picture of the complex and non-linear interrelations among the most important variables that impact on *WR* probability.

To complete this analysis we also run the same in- and out-of sample tests used to assess the RT model's accuracy. The results are in Table 3. The data in the table denote better in- and out-of-sample predictability for Logit 1. Specifically, for the in-sample the values for AUC, sensitivity and specificity are 0.7222, 0.4778, and 0.8660 for Logit 1, while Logit 2 shows the following values: 0.6996 (AUC), 0.7111 (sensitivity) and 0.6196 (specificity). For the out-of-sample, the diagnostics for Logit 1 are even closer to those of our model, while the RT slightly outperforms the Logit specification. In terms of AUC, the values for the two models are the same (0.8554) but the RT shows better predictability of *WR*s (sensitivity is 0.9167) compared with Logit 1 (sensitivity is 0.7778). By contrast, Logit 1 is better than RT at minimising false alarms, since the value for sensitivity is 0.8276 against 0.6149 of our EWS. We thus confirm the consistency between the RT and Logit specifications, but only when interactions are taken into account.

V.4 LTCM Vs. Sub-Prime Crises

To better inspect the two major systemic events we carried out RT analysis over the two sub-periods January 1998–December 1999 and January 2007–September 2008. In so doing, we expected to detect what really happened in both crises, clarifying the reasons why many funds experienced extreme negative returns.

Figures 2 and 3 report the risk stratification for the two sub-periods, while the diagnostics of model accuracy are reported in Table 4. The LTCM collapse seems to be a pure contagion event, since the higher risk is for those cases in which the proxy is greater than 3 ($C_t > 3$). Interestingly, the extreme risk cluster is for substantial leverage commonality ($\text{DCC Beta} > 0.3008$), and all cases clustered within such a node (node #5) are observations over the months August 1998–October 1998. This suggests that the main reason underlying the LTCM collapse was extreme commonality in leverage dynamics. In that period, the level in beta was substantial and thus high correlations were associated with a high leverage level. There is thus a significant difference between the LTCM and sub-prime crises.

In fact, the sub-prime crisis seems instead to be strongly linked to commonality in (filtered) returns. The risk partition obtained over the period January 2007–September 2008, reported in Figure 3, shows that the crowded trades and contagion proxies lead to an extreme risk cluster. In this cluster, where $\Pr(WR) = 0.8569$ (node #8) and $\text{DCC Filter} > 0.8293$, funds tend to crowd more and more as signalled by $d\text{DCC Filter}$, which is required to be stable or positive ($d\text{DCC Filter} > -0.0012$). Such a partition was mainly shown in the observations of July 2008 and September 2008, when the number of *WRs* was six and eight, respectively. Similarly, for August 2007, when the number of *WRs* was five, the model clustered the corresponding observations within a node with $\Pr(WR) = 0.5296$ (node #3) (characterised by slightly high crowded trades), but that was less than 0.8293, together with a substantial drop in leverage commonality ($d\text{DCC Beta} > -0.0476$) due to substantial de-leveraging in the summer 2007 as commented on above.

As a whole, contagion is clearly a leading indicator, since the probability estimates are for all clusters greater than 0.35 when more than one other fund style experiences a *WR*, except for funds that have moderate commonality in returns ($\text{DCC Filter} \le 0.7797$) and no extreme contagion ($C_t \le 4$), for which the probability estimate is $\Pr(WR) = 0.0708$ (node #4). These funds denote the medium and high values of probability to get extreme negative returns. Moreover, in times of no contagion ($C_t \le 1$), the risk arises for funds denoting a negative return commonality, i.e. for Strong Short Bias funds. For these funds, we have indeed a moderate risk profile with $\Pr(WR) = 0.2361$ (node #1).

From a pure statistical viewpoint, for both sub-periods the accuracy of the model is high, as proven by the diagnostics reported in Table 4, which documents the ability to correctly classify *WR* at 77.14% (1998–1999) and 80.56% (2007–09/2008) of the number of total cases, as well as for no-*WR* at 96.22%% (1998–1999) and 85.63% (2007–09/2008) of total tranquil time observations.

VI. Conclusion

This paper developed an EWS for hedge funds based on specific red flags that help stratify the risk of future extreme negative returns. To do this we relied on RT analysis through which the predictor space is partitioned by a series of splitting rules based on specific thresholds, which act as risk signals. What we find to be interesting and promising for future works is that such a technique is not vulnerable to the common criticisms of parametric approaches and allows us to uncover forms of non-linearity and complexities as well as 'regime shifts'. While contagion and clustered negative returns in the hedge fund industry are now well documented (Boyson et al., 2010), what we need to know is how to signal warnings to be used to prevent potential abnormalities that could propagate at a systemic level. The methodology proposed in this paper provides an effective and pragmatic answer to this question, realising an EWS based on a collection of binary rules of thumb such as "$x_{ji} \leq s_j$" and "$x_{ji} > s_j$" for each predictor *x,* thus realising a risk stratification approach that can capture situations of extreme risk whenever the value of the selected variables *x* exceeds the pre-specified threshold *s*.

Our empirical findings prove that contagion, leverage commonality, crowded trades and liquidity concerns are the leading indicators for monitoring the risk dynamics of hedge funds. We document that our EWS estimated using data from 1998 to 2006 would have been able to predict more than 90% of the total worst returns occurred over the period 2007–2008. We also checked the robustness of our model by comparing its results with those from a traditional Logit model in order to prove the consistency between the two when interaction terms are included in the Logit specification.

A closer look at the LTCM default and sub-prime crisis showed the changing nature of systemic risk. The LTCM collapse was mainly due to extreme interconnectedness in leverage dynamics but no crowded trades. By contrast, the sub-prime crisis was associated with crowded trades together with dramatic de-leveraging and de-correlations in leverage dynamics due to liquidity concerns together with strong crowded trades.

References

Adrian, T., 2007, Measuring Risk in the Hedge Fund Sector, *Federal Reserve Bank of New York Current Issues in Economics and Finance*, 3, 1–7.

Alexander, C., 2002, "Principal Component Models for Generating Large GARCH Covariance Matrices", *Economic Notes*, 2, 337–359.

Bekaert, G., Harvey, C. and A. Ng, 2005, "Market Integration and Contagion", *Journal of Business*, 78, 39–69.

Berben, R. P. and Jansen, W. J., 2005, "Comovement in International Equity Markets: A Sectoral View", *Journal of International Money and Finance*, 24, 832–857.

Billio, M., Getmansky, M., Lo, A. and L. Pelizzon, 2012, "Econometric Measures of Systemic Risk in the Finance and Insurance Sectors", *Journal of Financial Economics*, 104, 535–559.

Bollen, N., Whaley, R., 2010, "Hedge Fund Risk Dynamics: Implications for Performance Appraisal", *Journal of Finance*, 64, 985–1035.

Boyson, N., Stahel, C., Stulz, R., 2010, "Hedge Fund Contagion and Liquidity Shocks", *Journal of Finance*, 65, 1789–1816.

Breiman, L., J. Friedman, R. Olshen and C. Stone (1984). *Classification and Regression Trees*. Wadsworth Inc., California.

Brunnermeier M. and L. H. Pedersen, 2009, "Market Liquidity and Funding Liquidity", *Review of Financial Studies*, 22, 2201–2238.

Brunnermeier M. and S. Nagel, 2004, "Hedge Funds and the Technology Bubble", *Journal of Finance*, 59 (5), 2013–2040.

Buraschi, A., R. Kosowski, and F. Trojani, 2012, "When There Is No Place to Hide: Correlation Risk and the Cross-Section of Hedge Fund Returns", Working Paper.

Christodoulakis G. A. and S E. Satchell, 2002, "Correlated ARCH (CorrARCH): Modelling the Time-Varying Conditional Correlation Between Financial Asset Returns", *European Journal of Operational Research*, 139 (2002) 351–370

Duttagupta R. and P. Cashin, 2011, "Anatomy of Banking Crises in Developing and Emerging Market Countries", *Journal of International Money and Finance*, 30, 354–376.

Forbes, K. and R. Rigobon, 2002, "No Contagion, Only Interdependence: Measuring Stock Market Comovements", *Journal of Finance*, 57, 2223–2261.

Fung, W, and D. A. Hsieh, 2007a, "Will Hedge Funds Regress Towards Index-Like Products?", *Journal of Investment Management*, 2, 56–80.

Fung, W, and D. A. Hsieh, 2007b, “Hedge Fund Replication Strategies: Implications for Investors and Regulators”, *Banque de France Financial Stability Review*, 10, 55–66.

Fung, W. and D. Hsieh, 2001, “The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers”, *Review of Financial Studies*, 14, 313–341.

Fung, W. and D. Hsieh, 2004, “Hedge Fund Benchmarks: A Risk Based Approach”, *Financial Analyst Journal*, 60, 65–80.

Getmansky, M. , Lo, A. and I. Makarov, 2004, “An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns”, *Journal of Financial Economics*, 3, 529–609.

Lahrecha A. and K. Sylwester, 2011, “U.S. and Latin American Stock Market Linkages”, *Journal of International Money and Finance*, 30, 1341–1357.

Longin, F. and B. Solnik, 1995, “Is the Correlation in International Equity Returns Constant: 1960-1990?”, *Journal of International Money and Finance*, 14, 3–26.

Manasse, P. and N. Roubini (2009), “Rules of Thumb for Sovereign Debt Crises”, *Journal of International Economics*, 78, 192–205.

Morris, S. and H. Shin, 2004, “Liquidity Black Holes”, *Review of Finance*, 8, 1–18.

Pástor, L. and R. Stambaugh, 2003, “Liquidity Risk and Expected Returns”, *Journal of Political Economy*, 111, 642–685.

Patton, A. and T. Ramodarai, 2012, “On the Dynamics of Hedge Fund Risk Exposures”, *Journal of Finance*, forthcoming.

Savona, R. and M. Vezzoli, 2012, “Multidimensional Distance to Collapse Point and Sovereign Default Prediction”, *Intelligent Systems in Accounting Finance and Management*, 19, 205–228.

Savona, R., 2012, “Risk and Beta Anatomy in the Hedge Fund Industry”, *European Journal of Finance*, forthcoming.

Shleifer, A. and R. Vishny, 1997, “The Limits of Arbitrage”, *Journal of Finance*, 52, 35–55.
Stein, J., 2009, “Presidential Address: Sophisticated Investors and Market Efficiency”, *Journal of Finance*, 64, 1517–1548.

Vezzoli, M., Stone, C. (2007). “Cragging”. Book of Short Papers CLADAG 2007, University of Macerata, September 12–14, Macerata, Italy.

Table 1: In-Sample and Out-Of-Sample Model Accuracy – Worst Returns

	In-Sample 1998-2006	**Out-Of-Sample 2007-09/2008**
Brier Score	0.1412	0.3004
Optimal Cut-off	6.70%	8.30%
Youden Index	38.96%	53.16%
AUC	0.7294	0.7115
Numbers of *WR*	90	36
WR correctly classified	59	33
Sensitivity	65.56%	91.67%
Specificity	73.40%	61.49%

The table shows the diagnostics used to assess the models' accuracy in-sample (1998–2006) and out-of-sample (2007–09/2008), namely the Brier score computed using (16), the Optimal Cut-off, which is the probability value used to maximise the Youden index, obtained as $[(1-\alpha)+(1-\beta)-1]$ with α and β the type-I and type-II errors, respectively. AUC is the area under the ROC curve. The table also reports the overall number of worst returns (*WR*) and the number of worst returns correctly classified. Sensitivity and Specificity are computed as 1 minus type-I errors and 1 minus type-II errors, respectively.

Table 2: Logit Analysis on Worst Hedge Fund Returns – 1998-2006

Variable	**Logit 1**	**Logit 2**
Constant	**-3.101**	**-3.183**
	(0.000)	(0.000)
C_t	**0.318**	**0.419**
	(0.001)	(0.000)
$\text{DCC Beta}_{t\text{-}1}$	**-0.828**	-0.408
	(0.033)	(0.273)
$\text{DCC Filter}_{t\text{-}1}$	0.230	0.279
	(0.748)	(0.699)
$d\beta_{t\text{-}1}$	-0.588	-0.051
	(0.136)	(0.874)
$\text{AUM}_{t\text{-}1}$	**3.234**	**2.972**
	(0.003)	(0.005)
$\text{CS}_{t\text{-}1}$	**-190.415**	**-211.618**
	(0.035)	(0.006)
$(C_t \cdot \text{DCC Filter}_{t\text{-}1} \cdot \text{CS}_{t\text{-}1})$	27.562	-
	(0.910)	
$(C_t \cdot \text{DCC Beta}_{t\text{-}1})$	**1.126**	-
	(0.004)	
$(C_t \cdot \text{DCC Beta}_{t\text{-}1} \cdot d\beta_{t\text{-}1})$	**2.324**	-
	(0.002)	
Log likelihood	-271.155	-279.104
MSE	0.069	0.071
McFadden R-squared	0.120	0.094

The table reports the estimation results of the Logits over the period 1998–2006. Logit 1 includes the covariates selected by the RT and interaction terms as in equation (12). Logit 2 includes only the variables selected by RT without the interaction terms. Numbers in bold font are significant at the 0.1 level. P-values are in parentheses.

Table 3: In-Sample and Out-Of-Sample Model Accuracy – Logit

	Logit 1		Logit 2	
	In-Sample 1998-2006	Out-Of-Sample 2007-09/2008	In-Sample 1998-2006	Out-Of-Sample 2007-09/2008
Brier Score	0.1376	0.2193	0.1429	0.2247
Optimal Cut-off	12.00%	16.70%	6.70%	8.40%
Youden Index	34.38%	60.54%	33.07%	44.64%
AUC	0.7222	0.8554	0.6996	0.8234
Numbers of *WR*	90	36	90	36
WR correctly classified	43	28	64	32
Sensitivity	47.78%	77.78%	71.11%	88.89%
Specificity	86.60%	82.76%	61.96%	55.75%

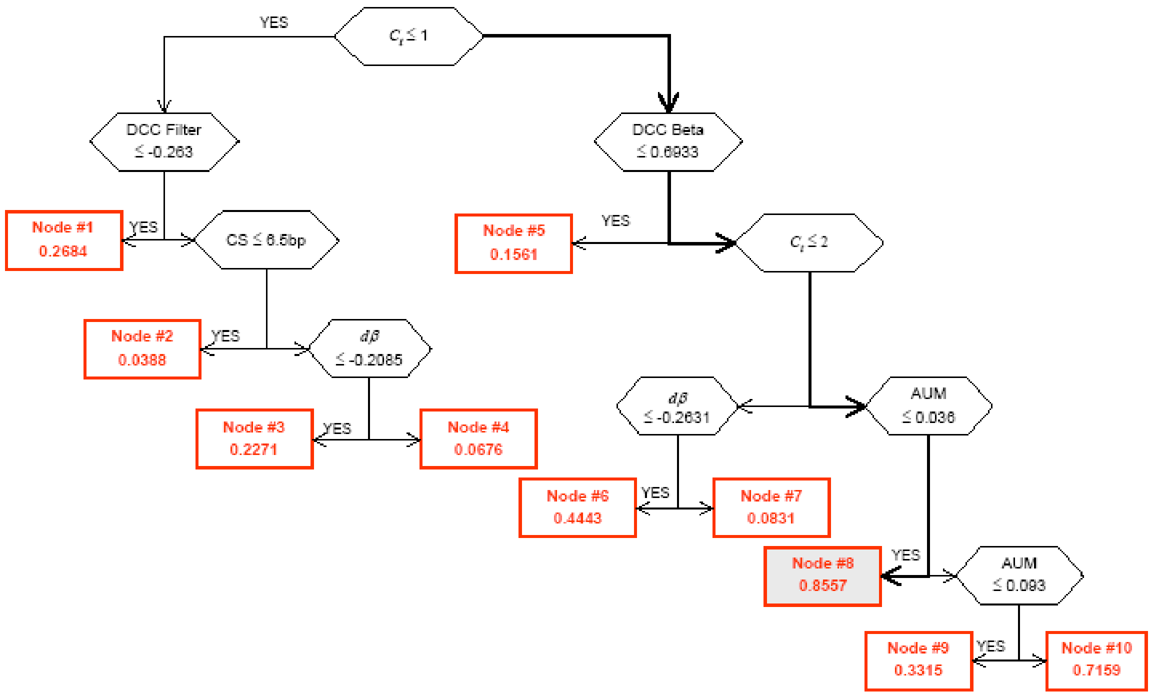
The table reports the same diagnostics used in Table 1 computed for the Logit specifications. Logit 1 includes the covariates selected by the RT and the interaction terms as in equation (12). Logit 2 includes only the variables selected by RT without the interaction terms.

Table 4: LTCM vs. Sub-Prime Crises Model Diagnostics – Worst Returns

	LTCM 1998-1999	Sub-Prime 2007-09/2008
Brier Score	0.1266	0.1796
Optimal Cut-off	50.00%	35.00%
Youden Index	73.36%	66.19%
AUC	0.8792	0.8836
Numbers of *WR*	35	36
WR correctly classified	27	29
Sensitivity	77.14%	80.56%
Specificity	96.22%	85.63%

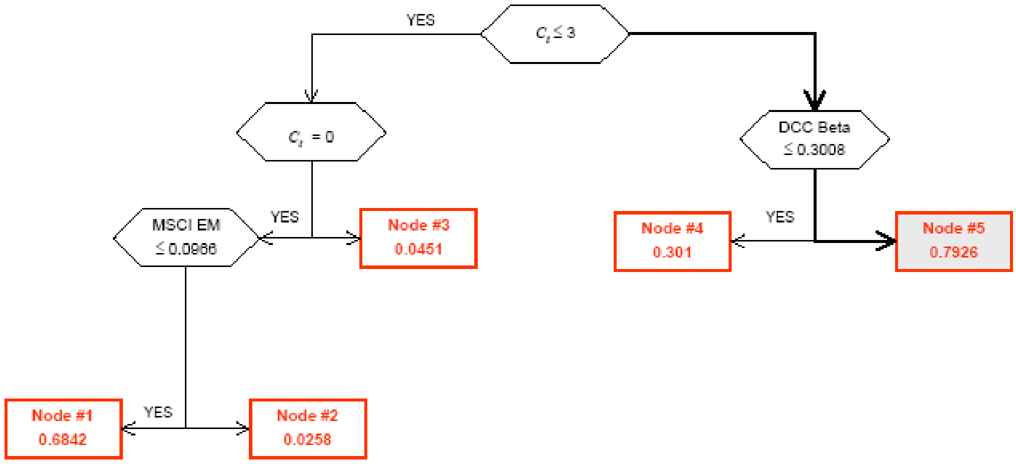
The table reports the same diagnostics used in Table 1 computed for the sub-periods 1998–1999 and 2007–09/2008.

Figure 1: EWS for Worst Returns – from January 1998 to December 2006



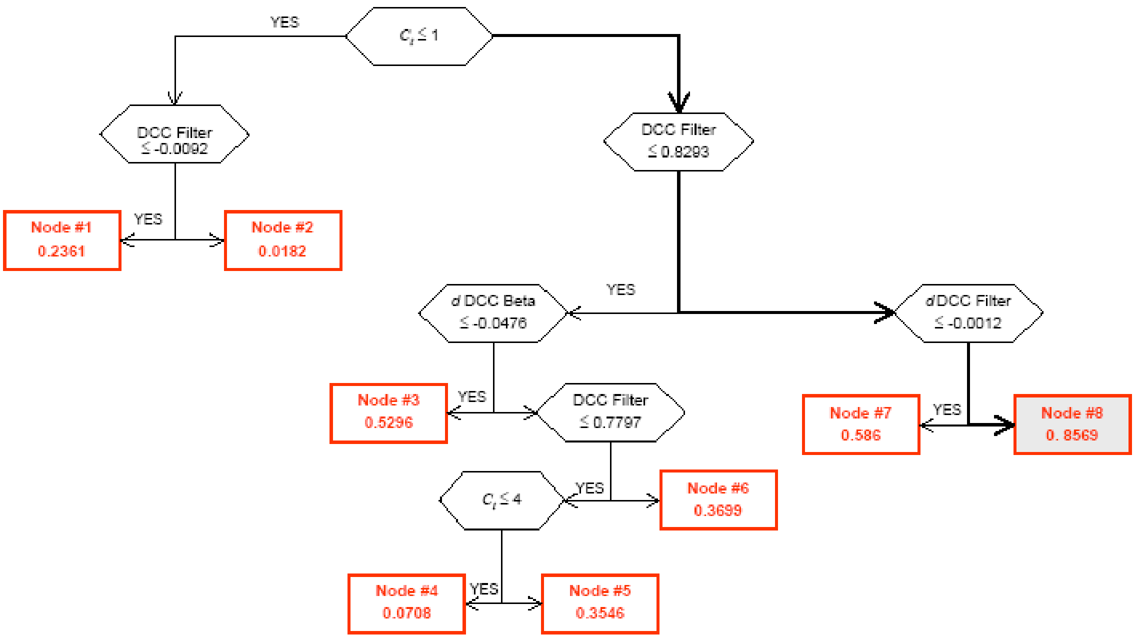
The figure depicts the structure of the EWS for worst returns (*WR*) estimated over the period 1998–2006. For each split, we specify the variable and corresponding threshold and indicate the paths towards the terminal nodes. The values reported within each terminal node are the estimated probabilities of *WR*. The most risky node is indicated by the grey area, which also highlights the paths towards the higher probability with the bold line.

Figure 2: EWS for Worst Returns – LTCM crisis (1998–1999)



The figure depicts the structure of the EWS for *WRs* estimated over the period 1998–1999 as in Figure 1.

Figure 3: EWS for Worst Returns – Sub-Prime crisis (01/2007–09/2008)



The figure depicts the structure of the EWS for *WRs* estimated over the period 2007–09/2008 as in Figure 1.

Appendix

Annex A presents the data used in the empirical analysis, while Annex B describes the regression tree strategy used to realize the Early Warning System for hedge funds. The comparative analysis between the regression tree algorithm used in our empirical analysis and alternative approaches are displayed in Annex C.

Annex A. Data description.

Table A1 reports summary statistics for CSFB/Tremont indexes (Panel A), 7+1 FH Risk Factors (Panel B), PRSs (Panel C), and the 3 liquidity factors of Pastor and Stambaugh (2003) over the periods 01/1998-12/2006 (in-sample) and 01/2007-08/2008 (out-of-sample). Mean is the monthly mean return. Min and Max are the minimum and maximum monthly return respectively. StdDev is the monthly standard deviation. Statistics for Credit spread in Panel B are expressed in basis points.

Table A1: Descriptive Statistics of Hedge Fund Indexes, 7+1 FH Risk Factors, PRSs and Pastor-Stambaugh Liquidity Factors

	Mean (%)		Min (%)		Max (%)		StdDev (%)	
	98-06	07-Sept08	98-06	07-Sept08	98-06	07-Sept08	98-06	07-Sept08
Panel A: Hedge Fund Index Returns								
Convertible Arbitrage	0.59	-0.74	-4.83	-12.26	3.40	2.18	1.42	3.19
Dedicated Short Bias	-0.29	0.55	-8.84	-7.30	22.55	10.31	5.20	4.79
Emerging Markets	0.61	-0.02	-23.19	-8.93	15.17	5.48	4.27	3.31
Equity Market Neutral	0.68	0.50	-1.01	-1.41	2.33	1.65	0.65	0.72
Event Driven	0.71	0.15	-11.93	-5.75	3.17	3.24	1.73	2.15
Fixed Income Arbitrage	0.30	-0.38	-7.11	-6.80	1.89	2.07	1.13	2.29
Global Macro	0.74	0.70	-11.70	-6.63	10.00	4.44	2.52	2.56
Long/Short Equity	0.86	-0.03	-11.59	-7.81	12.84	3.73	3.09	2.89
Managed Futures	0.52	0.65	-8.71	-4.79	9.79	6.61	3.46	3.72
Multi-Strategy	0.63	-0.16	-4.91	-7.35	2.61	3.02	1.06	2.22
Panel B: 7+1 FH Risk Factors								
S&P	0.54	-0.30	-14.58	-8.60	9.67	4.75	4.45	3.40
Size Spread	0.30	0.20	-16.37	-3.64	18.41	4.69	4.00	1.96
10yr Treasury Yield	-0.01	-0.03	-0.53	-0.38	0.65	0.35	0.22	0.20
Credit Spread (basis points)	0.03	-7.90	-48.00	-33.00	25.00	24.00	13.75	16.86
PTFSBD	-2.26	-1.71	-25.36	-15.84	68.86	17.38	14.87	9.55
PTFSFX	0.01	3.04	-30.00	-25.75	55.34	66.01	17.22	19.92
PTFSCOM	-0.74	2.50	-23.04	-24.20	64.75	39.62	13.60	16.87
MSCI EM Index	1.16	0.65	-29.29	-12.59	13.55	11.02	7.01	6.71
Panel C: PRSs								
VIX	20.96	20.36	10.91	10.42	44.28	30.24	6.95	5.13
Change in 3m Tbill	0.00	-0.18	-0.72	-0.86	0.45	0.44	0.20	0.33
Term Spread	1.49	1.05	-0.53	-0.38	3.70	2.56	1.25	1.03
Innovation in S&P Vol	0.46	2.92	-9.81	-6.61	23.65	35.93	5.83	9.35
Panel D: Pastor-Stambaugh Liquidity Factors								
PS1	-3.59	-5.03	-33.42	-28.72	19.96	8.34	8.05	10.00
PS2	-0.03	-1.16	-27.12	-17.66	28.62	12.01	7.56	6.96
PS3	1.14	0.96	-10.13	-6.61	10.99	7.14	3.59	3.14

Annex B. Regression Tree Algorithm.

In this annex we briefly describe the CRAGGING (CRoss-validation AGGregatING) algorithm through which we realized the EWS in our analysis. This is the novel regression tree-based algorithm introduced in Vezzoli and Stone (2007), and recently used in Savona and Vezzoli (2012), which was proposed to remove some of the problems of the regression tree analysis when dealing with panel data and other types of structured data. The traditional approach assumes indeed that the covariates are i.i.d. within each region (node) and independent across regions, when instead autocorrelations and other latent dependencies could play a major role especially in panel data.

Having a panel data (*Y*, **X**) with *N* observations for $j = 1,\ldots,j,\ldots,J$ units (in our case the hedge fund indexes) in each *t*, with $t = 1,\ldots,t,\ldots,T$ and then $J \cdot T = N$, we denote with $\mathcal{L} = \{1,\ldots,j,\ldots,J\}$ the set of units and with $x_{jt-1} = \left(x_{1jt-1},\ldots,x_{rjt-1},\ldots,x_{Rjt-1}\right)$ the vector of predictors of unit *j* observed at time $t-1$ where $j \in \mathcal{L}$. The algorithm proceeds in two main steps:

- In the first step, $\mathcal{L}$ is randomly partitioned into $\mathcal{L}_V$ test sets and one of these is taken out of the observations used for estimation and reserved for testing. The corresponding training (estimation) set, we denote by $\mathcal{L}_V^c = \mathcal{L} - \mathcal{L}_V$, is used for estimation by repeatedly removing one unit (hedge fund index) per time and testing the corresponding regression tree on the same test set. This type of perturbation, conceived with the end of maintaining the hierarchical structure of the panel data, is repeated for all the $\mathcal{L}_V$ test sets[***].

- In the second step, the vector of the CRAGGING predictions computed in every test set is used to replace the dependent variable *Y*, so growing a single tree then obtaining a parsimonious Final Regression Tree (FRT), with good predictions (accuracy), better interpretability and minimized instability.

For more technical details see Savona and Vezzoli (2012), Appendix A.

[***] In this way, each unit always appears either in training set used for estimation, or in the test set used for validation, but never in both at the same time.

Annex C. Comparative Analysis

For robustness check, we compared our EWS with two alternative regression tree approaches. The first is the original **CART** (Classification and Regression Trees) algorithm introduced in Breiman et al. (1984). The second is the more recent conditional inference trees procedure introduced in Hothorn et al. (2006) to cope with the overfitting and variable selection problems induced by recursive fitting procedure (the algorithm is implemented in the **Party** add-on package to the R system for statistical computing available at http://CRAN.R-project.org/). Using the same diagnostics computed in Section V.3.1, we focused on the two sub-periods January 1998–December 2006 and January 2007–September 2008 to perform the in-sample and out-of-sample analysis, respectively. The results are reported in table A2. The diagnostics used to assess the models' accuracy are the Brier score, the optimal cut-off, which is the probability value used to maximise the Youden index, obtained as $[(1-\alpha)+(1-\beta)-1]$ with α and β the type-I and type-II errors, respectively, the AUC which is the area under the ROC curve. The table also reports the overall number of worst returns (*WR*) with the number of worst returns correctly classified, and sensitivity and specificity which are computed as 1 minus type-I errors and 1 minus type-II errors, respectively.

The results reported in table A2 prove that, (i) out-of-sample, our EWS is better than CART and Party, (ii) in-sample, CART is marginally the best classifier, while EWS and Party show similar performance.

In more depth, out-of-sample the sensitivity computed using the optimal cut-off points through the Youden Index is 0.9167 for EWS while CART and Party show 0.4722 and 0.6389, respectively. AUC are 0.8554 for EWS, 0.7201 for Party, and 0.5413 for CART. On the other hand, competing approaches are better non-WR classifiers: the specificity is 0.8391 for CART, and 0.7816 for Party. By contrast, in-sample the CART algorithm is slightly better than our approach in terms of AUC, which, in turns, shows a value marginally greater than Party. The Brier score values confirm out-of-sample superiority of the EWS and in-sample superiority of CART.

Table A2: In-Sample and Out-Of-Sample Model Accuracy

	CART		Party		EWS	
	In-sample	**Out-of-sample**	**In-sample**	**Out-of-sample**	**In-sample**	**Out-of-sample**
Brier	0.1031	0.4291	0.1336	0.3019	0.1412	0.3004
Cut-off*	8.50%	79.90%	14.80%	25.80%	6.70%	8.30%
Youden	42.55%	31.13%	39.13%	42.05%	38.96%	53.16%
AUC	0.7372	0.5413	0.7153	0.7201	0.7294	0.8554
Nobs of WR	90	36	90	36	90	36
WR correcly classified	55	17	56	23	59	33
Sensitivity	61.11%	47.22%	62.22%	63.89%	65.56%	91.67%
Specificity	81.44%	83.91%	76.91%	78.16%	73.40%	61.49%