Decomposing systemic risk of the edge funds industry: An approach based on Extreme Value Theory;

Auteur : Hübner, Philippe
Promoteur(s) : Hambuckers, Julien
Faculté : HEC-Ecole de gestion de l'Université de Liège
Diplôme : Master en ingénieur de gestion, à finalité spécialisée en Financial Engineering
Année académique : 2020-2021
URI/URL : http://hdl.handle.net/2268.2/11690

Decomposing systemic risk of the hedge funds industry: An approach based on Extreme Value Theory

Jury :
Promoter :
Julien HAMBUCKERS
Readers :
Cédric HEUCHENNE
Pierrick CLERC

Dissertation by
Philippe HÜBNER
For a Master's degree in Business Engineering, specialization: Financial Engineering
Academic year: 2020/2021

Contents

1 Introduction

The goal of this research master's thesis is to provide a concrete and innovative insight about the contribution of the hedge funds industry to systemic risk. As defined by Adrian and Brunnermeier (2016), systemic risk should be understood as the risk that the stability of the whole financial system would be impaired, with potential negative impacts on the real economy. Though the hedge funds have been mainly studied from the point of view of their return and performance, the research community (but also investors and policy makers) is showing increasing interest in their high tail exposure but also their commonality, a contagion effect that could threat the financial system. This threat could find its root in the shadow banking phenomenon, a parallel capital market, highly unregulated, in which the hedge funds industry has been playing a key role. Though these actors are much less regulated than banks, many of them behave in a similar fashion and are therefore subject to the same risks, such as liquidity risk. Thus, it is not surprising to observe that some prominent organisations consider that they may represent a threat to the stability of the financial system, i.e. they display considerable systemic risk[1].

Starting from this isomorphism between the banking sector and the actors of the shadow banking world, we attempt in this thesis to adapt a model developed by van Oordt and Zhou (2019a, 2019b) and primarily suited for the study of the banking sector. Our work thus aims at quantifying and explaining the connection between the hedge funds and the global financial system. Moreover, this approach allows for a separation into two components of high relevance: the Pure Tail Risk (PTR), defined as the scale of one hedge fund's tail exposure compared to the system, and the Systemic Linkage (SL), defined as the tail dependence between this hedge fund and the system. This separation allows for the implementation of distinct methodologies that overcome the data challenges usually faced when dealing with hedge funds. These methodologies allow us to explain both measures separately and retrieve the net impact on the systemic risk measurement. The results that we uncover even show that some effects would not have been detected without this decomposition.

This work provides two main potential contributions. The first one consists in the reconciliation and aggregation of the two measures into one single metric. These have been already studied in the hedge funds sector but in a separate fashion, without simultaneously analyzing

[1]The focus and the recommendations of the Financial Stability Board and the G20 are good examples ("Implementation of G20/FSB Financial Reforms in Other Areas" (2019))

them in a single framework. To the best of our knowledge, our work is the first one to integrate the concept of high tail exposure of the industry and the dependence, in the extremes or in the whole distribution, of the funds' returns towards the system, intuitively linked with the commonality and the contagion effect. Our second contribution is the use of Extreme Value Theory (EVT) tools to overcome the specific data challenges when facing the hedge funds industry, for both the PTR and the SL. The LASSO-Generalized Pareto Regression developed by Chavez-Demoulin et al. (2016) and Hambuckers et al. (2018) allows for a parametric quantification of the PTR while providing its relationship with respect to hedge funds characteristics but also macro-economic variables. Concerning the SL, this work uses a parametric tail dependence estimation through the use of copulae, coupled with a regression tree and a Probit regression approach to overcome the data issues that are inherent to hedge funds (scarcity, low frequency of reporting) but also provide the explanatory power to this second component.

We believe that this approach has enabled us to obtain insightful outcomes in terms of explanation of the factors that drive or reduce the systemic risk. The results highlight the importance of incentive fee, the use of high-water mark and the size of individual funds as reducers of the systemic risk of hedge funds through both the PTR and the SL. They also stress the importance of liquidity exposures in the explanation and its positive correlation with the systemic risk, giving the first potential confirmation that liquidity exposures, shadow banking and systemic risk are well interconnected phenomenons. This confirms that the shadow banking phenomenon may put a real danger on the financial stability, but we also provide some hints regarding the 'how' of this potential threat. This unregulated market, and at least the hedge funds sector, must therefore be closely monitored in the future.

Thanks to the time varying feature of the systemic risk coefficient, mainly driven by the methodology underlying the PTR, this work shows that the systemic risk of hedge funds industry are way higher in times of stability than in periods of stress, such as the 2008 crisis or the 1998 LTCM crash. In complement to this previous observation, we also observe an anticipation effect before shocks, already covered by the work of Agarwal et al. (2017) and intuitively linked with the option-like behaviour of the industry, as studied by Fung and Hsieh (2004). Finally, the estimation shows that the systemic risk at the end of the period of study (2017) has never reached higher levels in the past, while at the same time the size of the industry has been hugely growing these two past decades. This stresses the urgency to apply a better regulatory monitoring of what is considered in this work as an unregulated and non-traditional banking system. Consequently, this work provides some useful clues for efficient potential actions but also provides insightful information on the dynamic of hedge funds extreme exposures.

2 Literature Review

To review the past literature of the hedge funds research, it can be first good to provide to the reader the basic principles when it comes to hedge funds performance, risk and returns. In a second part, the focus will be to draw an overall picture of the answers the research community has provided on the relationship between hedge funds, systemic risk, liquidity risk and shadow banking. This section can be seen as the literature to support the initial hypothesis of this work. On this basis, the last section will give a view on the different aspects of contribution this analysis provides and formalize its underlying hypothesis.

2.1 Hedge funds: a reminder

Before digging into the main subject of the research, it is important to remind the particularities of hedge funds and the different challenges these create.

Hedge funds are one of the main focuses for the scientific community for the very specific characteristics and exposures they carry. First, the strategies that hedge funds implement are very diverse and untraditional. The use of leverage through short positions, the arbitrage objective or the use of other sophisticated strategies are common practices and the basis of hedge funds' specificities. This results in very specific and nonlinear exposition towards equity market: an option-like behaviour as Fung and Hsieh (2004) demonstrated with their seven-factor model. Secondly, these strategies follow a highly dynamic strategic pattern, resulting in a potentially substantial evolution of the exposures over time. Agarwal et al. (2017) and Ben-David et al. (2011) even suggest that there might exist a genuine skill underlying this dynamic to anticipate market shocks, a result that we provide too regarding the tail exposures of the hedge funds. For example, hedge funds were able to limit their exposure to equity market and their tail risk before the 2008 crisis. This variability in the strategies and exposures can be explained through the macroeconomic context, such as indicators like the VIX (Billio et al., 2012), but also through intrinsic characteristics of the hedge funds, such as its past performance or the management fee (Aragon & Nanda, 2012). Thirdly, hedge funds carry an important tail risk, a high risk for extreme losses that would even be linked with higher returns according to Agarwal et al. (2009, 2017). Finally, there is an inherent lack of transparency in the industry. As the hedge funds investors are institutional or particular investors with specific knowledge, the targeted public is way more narrow. This results in a more permissive regulation, allowing for less disclosure[2]. Another reason stems from the strategical stake of these information as the real added-value of the hedge funds is their ”secret recipe” that underlies their strategies. This last aspect results in increased difficulties to gather and analyse data.

A last important aspect of the literature, which must be considered when analysing hedge funds risks and performance, such as our case, is the lack of reliability, and sometimes the biasedness, of the information that they provide. Though databases exist with reporting of different hedge funds across time and their key characteristics (management fees, size, strategy style), these can contain biases as highlighted by Fung and Hsieh (2002). These authors put emphasis on the survivorship bias and the backfill bias. Another data issue is the serial auto-correlation of the returns of hedge funds, as stressed by Getmansky et al. (2004), leading to a persistence in the reporting of hedge funds. According to Getmansky et al., one of the explanations is the incentives for managers to ”smooth” their returns across time, made possible through the exposures to particularly illiquid products, that may not have well-defined prices, leading to data manipulation opportunities. As a result, they can report lower value than expected to create a ”reserve” for bad times. This artificial persistence can lead to biases in risk measurement and performance analysis.

2.2 Shadow banking, systemic risk and liquidity risk

Besides their inherent specificities, hedge funds are increasingly studied for another reason: their stake in a parallel capital market also known as shadow banking. Acharya et al. (2012) describes the shadow banking as the fact that some intermediates that are not commercial or investment banks will however play the role, indirectly, of the banking sector. This is made possible through a market that enables banks to trade complex financial instruments they issue, such as CDO or CDS, in exchange of funds from these actors. As a result, risks revolving around those instruments created by banks are transferred outside the regulated financial system. We talk about liquidity exposure and credit risk as well as the transfer of the maturity transformation function banks usually perform, as pointed by Bengtsson (2016). However, one of the reason the banks are so intensively regulated is the systemic risk they carry, caused

[2] As a comparison, AIMD and UCITS directives impose much different sets of restrictions for alternative and mutual funds in the EU.

by the particularities of their exposures and activities. If these are transferred to other actors, we can expect that such systemic risk is transferred as well. The real problem behind this situation is that we have two players, banks and funds, that are carrying at least partially the same kind of risk but that are not equally regulated. Hedge funds are important stakeholders in this phenomenon as these are funds providers for banks in the relationship. According to Bengtsson (2016), their participation represented 47% of the CDO's market in 2007. This therefore results in an exposure to these specific risks. This aspect of shadow banking behind the hedge funds industry justifies the choice to use a model primarily designed and implemented for the systemic risk measurement in the banking sector.

Regarding the liquidity risk, it has been first considered by the work of Getmansky et al. (2004) with the serial auto-correlation of the hedge funds returns, as it was discussed earlier for its impact on the performance measurement. According to them, there are two justifications to assume that this measure is positively linked with liquidity risk. First, the intensity of auto-correlation is linked with the intensity of market frictions causing inefficiencies. Since the most important friction on the market is the lack of liquidity, this particularity should be in relation with liquidity exposures. The second possibility is the smoothing of the returns the managers perform, possible especially for particularly illiquid products, in order to provide a more advantageous picture of the fund's performance. This also results in an observed auto-correlation of fund returns. These two explanations are thus complementary and explain why the auto-correlation for hedge funds returns give a good proxy of the liquidity risk a particular fund carries. The conclusion of Teo (2011) about liquidity risk is the same: hedge funds bear higher liquidity risk than they should. This is the consequence of an asset and liability mismatch, which is typically a problem the banking sector faces. When comparing liquid and illiquid hedge funds, the author shows that the illiquid ones tend to have significant higher alpha, explaining the reason why hedge funds have strong incentives to be exposed to such risk. He highlights also the link with redemption period by pointing the fact that higher ones could generate incentives for managers to be exposed to higher liquidity risk. This aspect turns out to be significant too in our analysis and would explain why it has a positive correlation with systemic risk.

We have to ask an important question though: is this liquidity exposure really linked with systemic risk? As a starting point, as hedge funds have dynamic strategies and even sometimes market timing abilities, it might be possible to witness a common behaviour in particular contexts. This is what Racicot and Théoret (2016) show in their study of the common behaviour of hedge funds towards macroeconomic shocks. In parallel, Bussière et al. (2014) studied the commonality of hedge funds during the period end 2003 to 2006 through a PCA-analysis and evidenced an increase of it during this period. Through this PCA-analysis, they also emphasize the link between liquidity exposure and systemic risk. They build quantile-portfolio based on their principal component loading from the first PCA (the one explaining the most variance). This first PCA can be interpreted as the strength of the linkage or the commonality between the hedge funds. They derive the lag 1 auto-correlation coefficient of these portfolios based on the Getmansky et al. theory and evidence high illiquidity for high commonality quantiles. They conclude that one of the drivers of this commonality is the liquidity exposure borne by the whole hedge funds industry but also the exposure to emerging market. This last aspect turns out to be selected for the explanation of the PTR as a driver of the systemic risk in our model, which is consistent with these conclusions too. They also stress the potential danger that their activities carry on the whole financial system.

The conclusion of Hespeler and Loiacono (2017) is similar. Yet, their study of systemic risk relates to a risk within the hedge funds sector and not the potential effect on the financial system directly. Moreover, beyond a simple common behaviour, we can underline real contagion effect among hedge funds worst returns, such as in period of stress. This is what Boyson et al. (2010) show. According to these authors, this contagion could be linked with liquidity shocks, such as the financial crisis of 2008. In addition, Mhalla et al. (2020) study the tail dependence between hedge funds investment styles and observe, for some pairs, high dependence in times of crisis. This evidence involves that this contagion is present in the extremes but also that investment styles could be key variables for the explanation of the systemic risk of hedge funds.

This overall picture shows that we can conjecture the presence of a real link between the liquidity risk and the contagion effect, which is proven to exist even in the extremes of the distribution. The study of the threat the hedge funds sector until now has been mainly focusing on the commonality and the idea of contagion inside the industry. The approach we propose is thus, by the use of a model normally suited for the banking sector, to analyse the link between the sector and the system considered here as the whole financial market, now that we know there is a high commonality inside the hedge funds sector.

2.3 Contribution

From this overall picture, we can draw several observations and hypothesis. There is a systemic risk, mainly measured until then by an idea of contagion inside the sector. The literature on the subject and the shadow banking explanation seem to be the two tosses of the same coin. More globally, the literature has not yet been interested in a in-depth analysis of the determinants of such systemic risk in the hedge funds industry nor to consider the impact on the link with the real economy as one could do with the banking sector. We know that extrinsic and intrinsic variables can determine the exposures dynamic of hedge funds. While Bussière et al. (2014) set their focus on the overall commonality inside the hedge funds sector, this work focuses on overall sensitivity against extreme events, which encompasses an idea of commonality, but with the system and not within the industry.

The high tail exposure is also being widely studied. Recently, Gregoriou et al. (2021) studied the VaR dynamic of hedge funds with a non-linear method and showed the importance of the macro-economic uncertainty in the explanation of the downside risk exposures, while Bali et al. (2014) or Racicot and Théoret (2016) focused on the whole distribution of the returns. This work contributes first by aggregating the two aspects, the idea of commonality and the tail exposure, respectively through the Systemic Linkage (SL) and the Pure Tail Risk (PTR). Furthermore, as Gregoriou et al. (2021) prove that there is a link between the macro-economic uncertainty and the higher moments of the hedge funds returns, our work confirms this aspect and completes it by incorporating intrinsic factors to hedge funds into the explanation of the PTR. Besides Bussière et al. (2014) and Mhalla et al. (2020) show that extrinsic and intrinsic variables are significant in the explanation of the commonality, respectively in the overall distribution or in the extremes (through tail dependence study). Our work therefore manages to provide the same insight for the SL.

Additionally, many tools allowing for the analysis of systemic risk are generally suited for the case of the banking sector, where reporting frequency is much higher and available data are much more important compared to hedge funds. Our work also attempts to adapt those kinds of tools in order to face the challenges hedge funds represent when dealing with the implementation of such models.

As a result, provided that its reliability can be confirmed, an analysis of this kind could represent an interesting add-on to the tool kit of both regulators and financial actors in their efforts to better monitor and understand the systemic risk of hedge funds. By doing so, it would provide a new brick in the wall that shapes the answer to the challenge that represents an unregulated market with such systemic risk. In addition, a common key element between shadow banking and hedge funds systemic risk is the liquidity exposure. The shadow banking might explain its appearance and the literature tends to confirm its presence in hedge funds. If we are able to confirm that there is a link between this liquidity exposure and the systemic risk of hedge funds, we can get closer to the confirmation of this hypothesis. In order to reach this objective, the underlying core hypothesis of our work is the capability of hedge funds characteristics and external macro-economic variables to explain the tail exposure but also the extreme dependence with respect to the whole financial system.

3 Models and Methodology

3.1 The van Oordt - Zhou model for systemic risk

The model that is used for the analysis is based on the work of van Oordt and Zhou (2019a, 2019b). Their study covers the systemic risk in the banking system. The rationale of this model is to measure systemic risk as the sensitivity to shocks of the whole system. Therefore, the assumption is that a system shock can be approximated through the shock of a wide equity investment, such as the S&P500. To do so, they estimate a systemic risk coefficient with respect to a given equity investment that follows the equation

$$R_t^i = \beta_t^i R_t^s + \epsilon_t^i \quad where\ R_t^s < -VaR_t^s(\bar{p}). \tag{1}$$

In this equation, $VaR_t^s(\bar{p})$ denotes the value at risk (VaR) of the equity market at level $\bar{p}$. R_t^i and R_t^s are respectively the returns at time t of the hedge fund i and the returns at time t of an equity investment in the financial system, given that the given hedge fund return is below the so-called VaR of the equity investment, involving them to be negative. Finally, β_t^i is the following systemic risk, linking extreme shocks from the equity market to the hedge funds returns.

Through Extreme Value Theory, it has been shown by van Oordt and Zhou (2019a) that β_t^i can be approximated the following way:

$$\beta_t^i \approx \lim_{p \to 0} \tau_i(p)^{1/\zeta_t^s} \frac{VaR_t^i(p)}{VaR_t^s(p)} \tag{2}$$

where $\tau_i(p) = Pr(R_i < -VaR_i(p) | R_s < -VaR_s(p))$.

In this estimation, $\tau_i(p)$ is the tail dependence that we can describe as the probability for the hedge funds return to witness a high negative shock if the equity investment experiences the same type of shock. In this context, ζ_t^s is the tail index of the equity index returns at time t.

An advantage of this model is that it enables for a separation of the effect of the tail exposure (in a way scaled for every VaR of hedge funds by the same VaR of the equity investment) and the systemic linkage, which is related to a pure contagion effect. This is made through a log-transformation:

$$\log(\beta_t^i) \approx \log\left(\lim_{p\to 0} \tau_i(p)^{1/\zeta_t^s}\right) + \log\left(\lim_{p\to 0} \frac{VaR_t^i(p)}{VaR_t^s(p)}\right) \tag{3}$$

$$\log(\beta_t^i) \approx \log\left(\text{Systemic Linkage}\right) + \log\left(\text{Pure Tail risk}\right) \tag{4}$$

In their methodology, van Oordt and Zhou estimate the systemic risk coefficients through a non parametric approach for the VaR and systemic linkage. Once these coefficients are estimated on a rolling window basis, they regress these against explanatory variables intrinsic to banks.

3.2 Specificities of the hedge funds

Compared to the methodology the authors implement for the estimation of the aforementioned model, slight changes are needed in the case of hedge funds. The reporting period is less frequent, as banks report on daily-basis while hedge funds disclose on monthly-basis. In addition, the population of study is smaller and the disclosure less regulated, resulting in bias. Finally, the strategies adopted by the hedge funds are highly dynamic. This is why the methodology adaptation will be focused on providing a time-varying model (dynamic strategies) based on Extreme Value Theory (scarcity of data) while mitigating at the maximum the different bias highlighted in the literature. Though the model of van Oordt and Zhou incorporates this need to have time varying estimates through the use of rolling windows, this solution cannot be applied in our context as the size of the sample is not sufficient. We cope with these specific challenges differently for the Pure Tail Risk and the Systemic Linkage. For the Pure Tail Risk, we used a LASSO-Generalized Pareto Regression, using time-varying extrinsic variables and intrinsic characteristics of the individual hedge funds. For the Systemic Linkage, the approach consists in fitting copula distribution between the equity returns and one particular hedge fund or a group of hedge funds. Nevertheless, this approach does not work for specific hedge funds for which the tail dependence is estimated to be zero. For these cases, two approaches are considered: a regression tree and a Probit regression.

3.3 Estimation methods

Several esimtation procedures are needed for the aggregation into the systemic risk coefficient. We develop the way each component is estimated in the following sections. Our choices can be summarized as followed:

- Pure Tail Risk
  - Value at Risk of hedge funds ($VaR_t^i(p)$): LASSO-Generalized Pareto Regression (hereafter LASSO-GP Regression) for a parametric estimation that will simultanously cope with the scarcity of data and provide the explanatory power of the approach
  - Value at Risk of the equity index ($VaR_t^s(p)$): EVT-GARCH Model, which takes into account the heteroscedasticity of the equity index and its heavy taildeness.
- Systemic Linkage
  - Equity tail index (ζ_t^s): Generalized Pareto distribution fit on the monthly returns and use of the shape parameter ξ^s, assumed constant over time, where $\zeta_t^s = 1/\xi^s$. This enables us to use the monthly returns, avoiding us the problem of transformation from a daily tail index to a monthly tail index.
  - Tail dependence ($\tau_i(p)$) : copula distribution fit on grouped data for a parametric estimation of the tail dependence, as the non-parametric approach does not generate appreciable results.

3.3.1 LASSO-Generalized Pareto (GP) Regression

To estimate $VaR_t^i(p)$, we used a Generalized Pareto Regression such as introduced by Chavez-Demoulin et al. (2016), based on explanatory variables related to intrinsic and extrinsic characteristics of the individual hedge funds. With this method, we can estimate ξ_t^i and σ_t^i, respectively the shape and scale parameter of the distribution, that are explicitly time-varying thanks to the regression. This approach enables to treat at the same time the measurement and the explanation but also to implement a penalization procedure for the variables selection. To do so, we assume that:

$$-R_t^i \mid (-R_t^i > u) \sim GPD(\xi(x_t^i), \sigma(x_t^i), u) \tag{5}$$

where u is the location parameter and R_t^i corresponds to the returns of hedge fund i at time t. The Generalized Pareto Distribution (GPD) is given by

$$GPD(x; \xi, \sigma, u) = \begin{cases} 1 - \left(1 + \xi \frac{x-u}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - e^{-(x-u)/\sigma}, & \xi = 0. \end{cases} \tag{6}$$

In our context, the explanatory variables for ξ and σ are a subset of the available variables given by the dataset, but this subset can be different between the two variables. Therefore we can write:

$$-R_t^i \mid (-R_t^i > u) \sim GPD(\xi(x_{i,t}^{\xi}), \sigma(x_{i,t}^{\sigma}), u) \tag{7}$$

where $x_{i,t}^{\sigma}$ (resp. $x_{i,t}^{\xi}$) are the explanatory variables for σ (resp. ξ). We also have to constraint σ with $\sigma \geq 0$, which is not the case for ξ. This is why we assume

$$\log(\sigma(x_{i,t}^{\sigma})) = \alpha_0^{\sigma} + \sum_{l=1}^{p_\sigma} \alpha_l^{\sigma} x_{i,t}^{\xi}(l)$$
$$\xi(x_{i,t}^{\xi}) = \alpha_0^{\xi} + \sum_{l=1}^{p_\xi} \alpha_l^{\xi} x_{i,t}^{\xi}(l)$$

where l denotes the l^{th} element of the covariates $x_{i,t}^{\xi}$ and $x_{i,t}^{\sigma}$ while p_ξ and p_σ are the total number of covariates respectively for ξ and σ.

Under these assumptions, the estimation is done through Maximum Likelihood Estimation and we thus obtain a log-likelihood function that is

$$\mathcal{L}(u; \Theta; x_{i,t}^{\xi}; x_{i,t}^{\sigma}) = \sum_{t=1}^{T} \sum_{i=1}^{n_t} \log(gpd(\xi(x_{i,t}^{\xi}), \sigma(x_{i,t}^{\sigma}), u)) \tag{8}$$

where Θ is the set of parameters estimated through the linear equation above and n_t is the number of observations above the threshold at time t. An estimator of these parameters is obtained by maximizing equation (8) with respect to Θ such that

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(u; \Theta; x^{\xi}_{i,t}; x^{\sigma}_{i,t}). \tag{9}$$

From these different estimates we can derive the VaR that is provided by

$$\widehat{VaR}^{i}_{t}(p) = u + \frac{\hat{\sigma}^{i}_{t}}{\hat{\xi}^{i}_{t}} \left(\left[\frac{1-p}{\mathbb{P}(-R^{i}_{t} > u)} \right]^{-\hat{\xi}^{i}_{t}} - 1 \right) \tag{10}$$

where p is the level of the VaR. As equation (2) gives an approximation of the systemic risk coefficient with the convergence of p to zero, we have to set the level as low as possible, which has been to chosen to be 1%.

However, as the number of candidates in the explanatory variables set is relatively important, a selection process of these variables has to be performed based on the LASSO approach developed by Hambuckers et al. (2018). In this framework, we use a penalized log-likelihood function defined as

$$\mathcal{L}_{pen}((u; \Theta; x^{i}_{\xi,t}; x^{i}_{\sigma,t}) = \mathcal{L}(u; \Theta; x^{i}_{\xi,t}; x^{i}_{\sigma,t}) - n\mathcal{P}_{\nu}(\Theta) \tag{11}$$

where $n = \sum_{t=1}^{T} n_t$ is the total sample size present above the threshold u and $\mathcal{P}_{\nu}(\Theta)$ is the penalty with parameters $\nu = (\nu_{\sigma}, \nu_{\xi})$ and defined as

$$P_{\nu}(\Theta) = \nu_{\sigma} \sum_{l=1}^{p_{\sigma}} a^{\sigma}_{l} |\theta^{\sigma}_{l}| + \nu_{\xi} \sum_{l=1}^{p_{\xi}} a^{\xi}_{l} |\theta^{\xi}_{l}| \tag{12}$$

where θ^{σ}_{l} (resp. θ^{ξ}_{l}) is the l^{th} parameter associated to the equation of σ (resp. ξ). In this work, we use the classical LASSO involving that $a^{\sigma}_{l} = a^{\xi}_{l} = 1$.

As a result, this procedure requires to estimate the penalty coefficients ν_{σ} and ν_{ξ}. As explained by Hambuckers et al., no analytical solutions exist for solving the penalized likelihood function. It should then be solved numerically. Therefore, to estimate these to coefficients, a grid of possible combinations of the two values is created and evaluated based on the BIC criterion.

As a result, we obtain the set of estimated parameters through

$$\hat{\Theta}_{pen} = \arg\max_{\Theta} \mathcal{L}_{pen}(u; \Theta; x^{\xi}_{i,t}; x^{\sigma}_{i,t}). \tag{13}$$

3.3.2 Equity index model

Regarding the estimation of the VaR of the equity investment, we used a EVT-GARCH model developed by McNeil and Frey (2000) that is briefly summarized here. Let ($X_t, t \in \mathbb{Z}$) be a strictly stationary time series of returns and let assume that the dynamic of X is given by

$$X_t = \mu_t + \sigma_t Z_t, \tag{14}$$

where Z_t is a strict white noise process with mean zero and unit variance. In this framework, the principle is first to fit a GARCH-type model on the index returns and, from the estimated μ_t and σ_t, retrieve the residuals. Assuming these to be strict white noises, we can use the EVT to estimate a quantile of level p, denoted z_p through the fit of a Generalized Pareto Distribution on these residuals, in a similar way to equation (10) without the time varying fashion, as the white noise is assumed to be independent of t. As a result, we obtain

$$\widehat{VaR}^{s}_{t}(p) = \mu^s_t + \sigma^s_t z_p. \tag{15}$$

This approach allows to take into account the fat tails of the index returns distribution and the stochastic nature of its volatility with an heteroscedasticity assumption. As the hedge funds reporting frequency is on monthly basis, and therefore the corresponding VaR has also a time horizon of one month, we need also the VaR of the equity index to be on a monthly basis. Using the time series of monthly returns of this index would create some difficulty as the number of data beyond the threshold for the EVT analysis would not be satisfying. This is why it has been decided to perform this analysis on the daily returns time series and aggregate all the daily VaR of one single month into one single monthly VaR.

3.3.3 Systemic Linkage

In the context of this analysis, the Systemic Linkage represents the biggest challenge. Due to the scarcity of the data, already often mentioned, non-parametric methods for the tail dependence between the index and the hedge funds, as the ones developed by Embrechts et al. (1999) and used for the work of van Oordt and Zhou, are not possible. The estimation is thus based on a parametric estimation through the use of copula distributions. For completeness, a copula C links a multivariate (in our case bivariate) distribution function to its unidimensional marginal distribution functions. For the bivariate case, if we have two continuous random vector (X_1,X_2) with F the joint cumulative distribution function (cdf) and respective marginal cdf F_j for $j = 1, 2$, then the copula is the cdf of (U_1,U_2) where $U_j := F_j(X_j)$, assumed to be continuous with $U_j \sim \mathcal{U}(0, 1)$, through the application of the probability integral transform. Therefore we can denote

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \tag{16}$$

When relying on the work of Joe (1997), the lower tail dependence parameter, λ_L, of the random vector (X_1,X_2) can be defined as

$$\lambda_L = \lim_{u \to 0^+} \frac{C(u, u)}{u}. \tag{17}$$

In this framework we assume that $\tau_i(p) = \lambda_{L,i}$, which would be computed for the i^{th} hedge fund and be constant during its whole life. We thus need to provide an estimation of the marginal distribution of the hedge funds returns and the equity index, focusing more particularly on the lower tail of the distribution. For the equity index, we need to fit a distribution on monthly returns. Therefore, we cannot use the approach we used in the computation of the VaR of the equity index. Consequently, it has been decided to use only the GARCH approach to take into account the stochastic volatility with a assumption of normality. For the hedge funds, we used the information retrieved from the analysis of the LASSO-GP Regression by using the cdf of the GPD with the calculated σ and ξ while using the empirical cdf for the data that are not beyond the threshold.

Finally, several types of copula exist with or without implied lower or upper tail dependence. The choice of candidates depended on this last characteristic (for example, Gaussian copula was excluded) and the available tools provided by Matlab (software used for the implementation of the model) within its toolbox, through the *copulafit* function. Therefore, the Student, Gumbel and Clayton copula have been fitted and evaluated based on the BIC and AIC. However, using copula distribution does not entirely solve the problem: several tail dependence coefficients remain at a level of zero. This problem could be due, notably, to the 'GARCH-only' approach used for the equity index, though the fit of the copula depends on the whole distribution, and not only the extremes. This is why the methodology needs to be extended to encompass this problem.

The idea is to create groups of individual hedge funds first based on the investment style they carry. Mhalla et al. (2020) investigated this approach on their analysis on the connectedness between hedge funds investment style. However, as shown by the study of Bussière et al. (2014) on the commonality of hedge funds, other variables than only the investment style could discriminate the common behaviour of hedge funds. This is the reason why we used as a first approach a regression tree analysis with a subset of the explanatory variables used in the analysis of the VaR of the hedge funds. This subset corresponds to the variables that are (i) intrinsic to hedge funds and (ii) kept constant during the life of the hedge funds. To fit the regression tree, we use the lower tail dependence as the response variable used would be the lower tail dependence, calculated on hedge funds at the individual level. We allocate, on a random basis, data on a test and train dataset to determine the depth of the tree. As a second approach, we used a Probit regression using each hedge fund as one observation, for which the response variable is the calculated tail dependence and the explanatory variables the average of each candidate variables observed during the life of the fund. The two approaches have pros and cons. On the one side, the regression tree enables to take advantage of the aggregation of data by fitting a new copula on it. The other positive aspect is its variable selection process that we do not have to manage as we only have to control the right depth of the tree. However, the final tree tends to be relatively deep, involving limited interpretation opportunities. On the other side, the Probit model provides at the same time the estimation and a clear explanatory power, easier to interpret. Nevertheless, it does not take advantage of a new copula on an aggregation of data, though it provides instead unique lower tail dependence values for each hedge funds we have to manage. At the end, the results of the Probit model are chosen. Since the number aforementioned zero cases was marginal, the final results were confirmed to be robust to the choice between the two models. Figure 1 gives a schematic summary of the respective approaches adopted to estimate the different components and their respective place inside the equation (2).

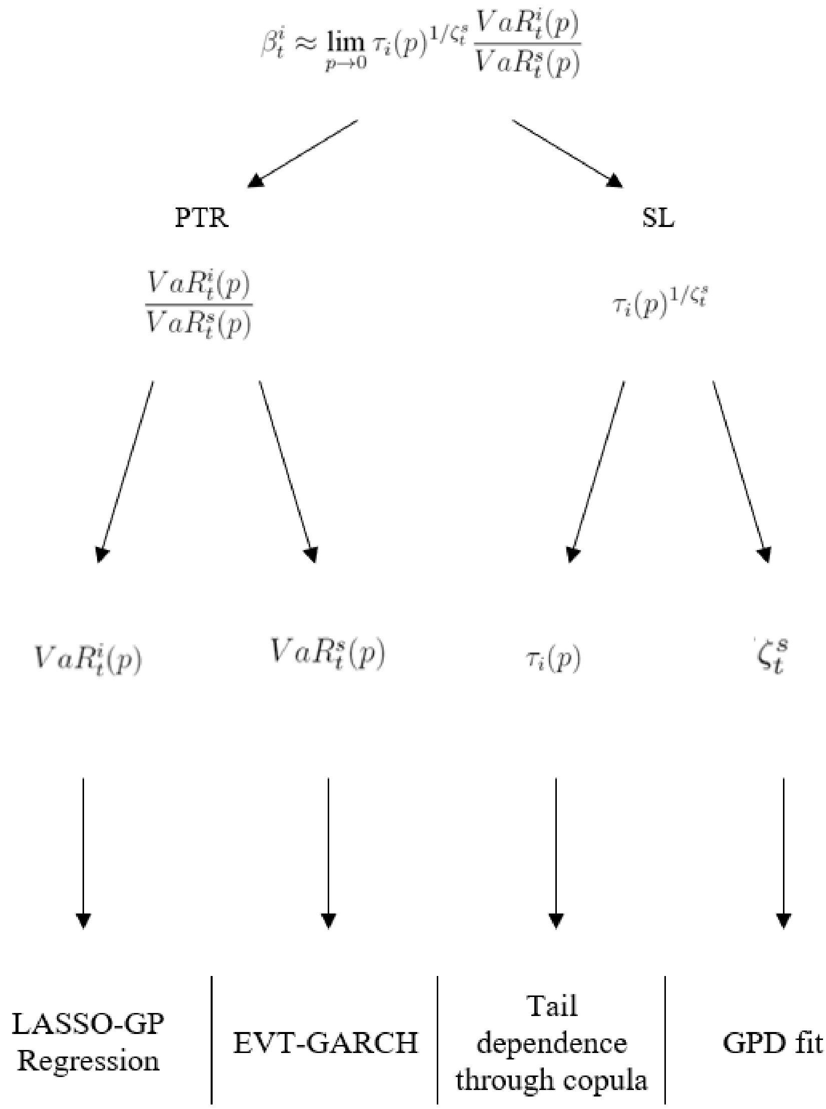

Figure 1: Summary of the methodology

4 Data description and treatment

4.1 Data description

4.1.1 HFR database

For the analysis, we use the Hedge Funds Research (HFR) database with multiple funds information and characteristics. The time interval on which we focus starts in February 1990 and ends in May 2017. It counts 897,084 monthly observations reported by 10,262 individual funds. In addition to the information provided by the HFR, extrinsic data are also provided (for example the VIX, or the MSCI returns).

For the purpose of the analysis, several characteristics of the database and the information it provides on the true population it targets are important to bear in mind. First, the USA is the most important country in which the hedge fund industry is implemented, as it can be clearly understood with Figure 2. This country is mentioned in around 65% of the observations and represent 64% of the individual hedge funds.

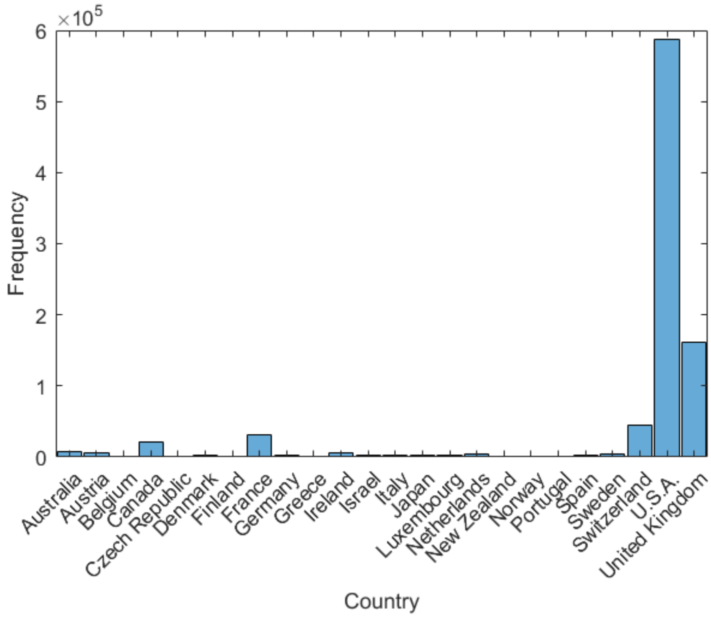

Figure 2: Country repartition, n=10262

What is also important to bear in mind is the dynamic of the population across time. Figure 3 gives an overview of this dynamic. The fact that we have both dead and alive funds that are present in the database enables it to be survivorship bias-free. In addition, we can also see that the sum of all the assets under management (AUM) reported at a given time has been hugely increasing until 2008 to be reduced afterwards due to the crisis of the same year. This increase is mainly driven by the increasing number of funds and therefore, at least partially (as the representativity of the HFR database can be questioned), by the increasing number of existing hedge funds. The exponential growth of the sector is not a secret though. What is interesting is the important shift of total AUM in 2008 compared to the one we can observe in terms of number of hedge funds reporting to the HFR. As the latter is less pronounced than the AUM shock, we could infer that the hedge funds that were stopping their activities due to a extreme shock tended to be big in terms of AUM, or at least in periods of stress. However, after 2008, the average AUM of the reporting funds increased while the number of hedge funds that were reporting decreased.

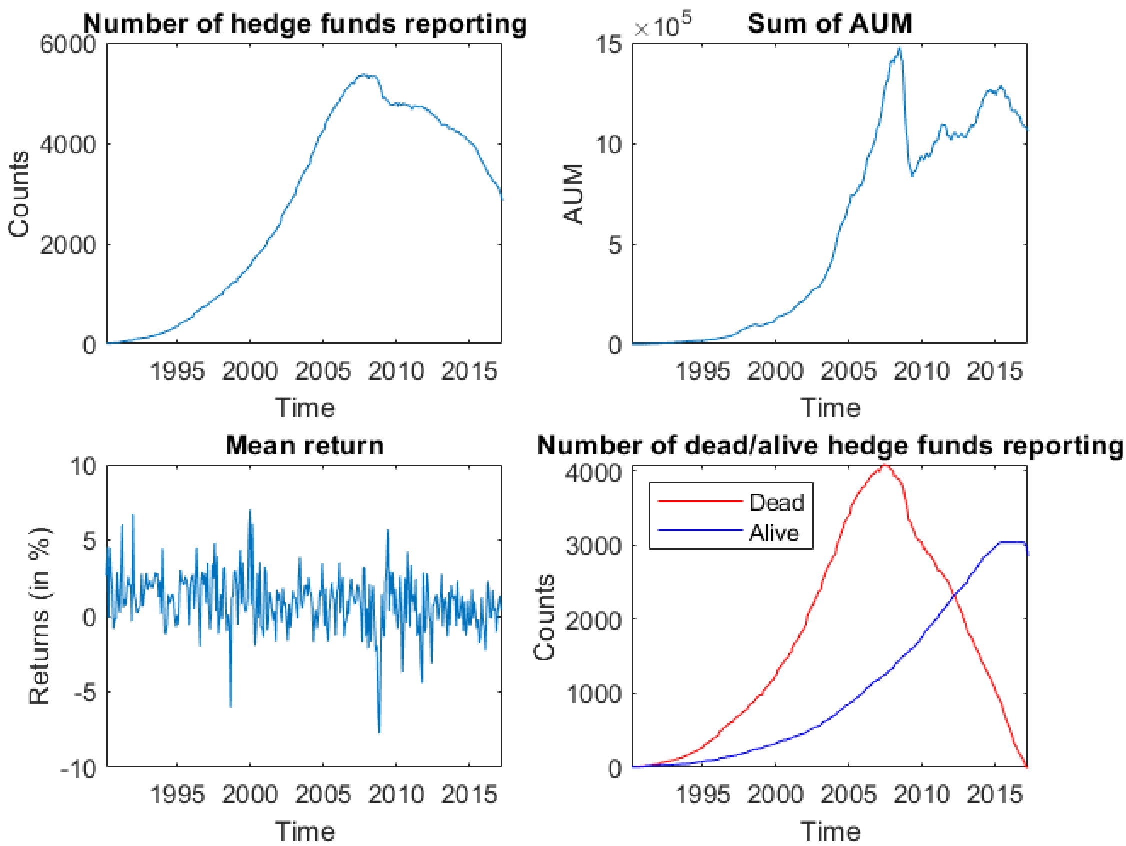

Figure 3: Evolution of the different characteristics of the population of hedge funds represented by the HFR database

4.1.2 Explanatory variables

Table 1 gives a statistical description of the numerical data we use for the analysis. Among the whole dataset, we provide candidate variables that could be intrinsic to one hedge fund, such as the management fee, the incentive fee, the use of leverage or the strategy each individual hedge fund pursues. This last variable is significant in our analysis and counts 5 different values:

- Equity Hedge: the objective of such a strategy is to provide equity-like returns with long and short positions on equity or equity-related products while reducing the downside risk of their investments
- Macro: the strategy revolves mainly around the attempt to profit from political and economic events through the movement of macro-economic variables
- Relative Value: the concept is to perform arbitrage by playing on the correction of pricing
- Event Driven: these funds are focused on corporate transactions such as merging or restructuring of companies
- Funds of Funds: these are funds that aggregate their investment in multiple hedge funds

Table 1: Descriptive statistics of the numerical variables

	Mean	Standard deviation	Median	Min	Max
Returns	0.59	4.48	0.60	-97.72	100.00
Assets	204.22	1139.32	40.70	0.00	82000.00
Management Fee	1.42	0.52	1.50	0.00	6.00
Incentive Fee	15.50	7.64	20.00	0.00	75.00
Lockup	106.94	200.53	0.00	0.00	2920.00
Redemptions	66.59	73.71	30.00	1.00	540.00
Advance Notice	40.81	32.44	30.00	0.00	365.00
CPI	0.00	0.00	0.00	-0.02	0.03
Ind. Prod.	0.00	0.01	0.00	-0.24	0.33
Unemployment Rate	6.26	1.95	5.60	1.80	27.90
MSCI	-0.40	4.43	-1.06	-10.91	19.05
EPU	110.04	77.72	87.58	10.12	626.03

Figure 4 shows that Equity Hedge hedge funds account for the majority of the population. We can also see that the large decrease in reporting number is driven by the funds of hedge funds, which particularly suffered from the crisis as stressed by Edelman et al. (2012), leading to an important number of liquidations.

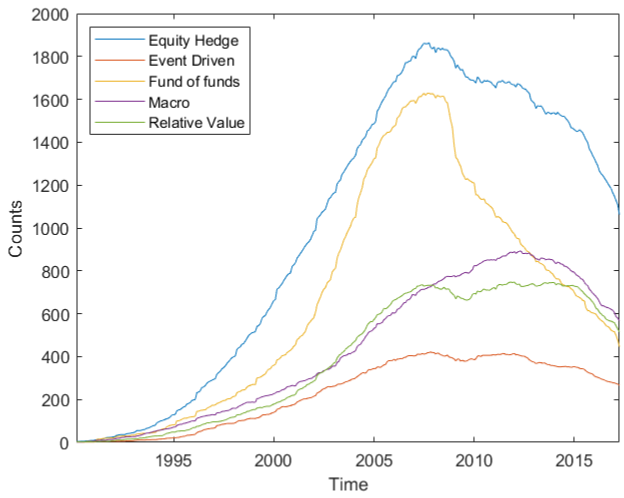

Figure 4: Number of reporting hedge funds by main strategy across time

There are also extrinsic variables that are related to macro-economic contexts such as the VIX, the MSCI or the EPU. These are thus classical variables.

An in-depth description of the variables is available in Appendix 7.1.

4.1.3 The equity index

To model the systemic risk of the banking sector, van Oordt and Zhou use an equity index as proxy of the system. The approach of our analysis is the same and our choice is to use the S&P500. The justification of such decision is twofold. First, the model of Fung and Hsieh (2004) uses this index for equity market factor they provide in their seven-factor model. Secondly, as it was pointed out in the country repartition of the hedge funds, most of them are implemented in the USA, making the S&P a more relevant choice than other equity indexes. The idea here is to find a balance between a large enough and a sufficiently linked proxy of the system.

4.2 Specific data treatment

4.2.1 Information bias in the HFR database

Several other data providers exist (Lipper TASS, BarclayHedge, EurekaHedge) and several authors emphasized the potential insight the researchers can gain in the performance analysis of the funds by merging these together. However, according to Joenväärä et al. (2019), the HFR database is the most suited one if a researcher has access to only one data provider as it is free of survivorship bias. Another often treated bias is the backfill bias, closely related to survivorship bias. It involves that the performance of hedge funds are overestimated between their inception and the date at which they are selected to be part of the database, as they tend to postpone bad performance to a later date where their performance is better. As explained by Fung and Hsieh (2009), the most common approach to deal with this bias is to select only hedge funds whose inception is relatively close to the moment they are selected to be part of the database. However, the authors drew the conclusion that such approach could be too costly in terms of information, as these are partially destroyed. In our context, we have already data free of survivorship bias. More importantly, the scarcity of data is key when it comes to EVT estimation. We thus assume that the justification of Fung and Hsieh (2009) is sufficient to put aside this consideration.

4.2.2 Liquidity model of Getmansky et al. (2004)

Though we have a relatively large set of potential variables for the explanation of the systemic risk, none of them are directly linked with a potential liquidity exposure. To fill this gap, we refer to the model of Getmansky et al. (2004) that studies the measurement of the illiquidity in hedge funds. To do so, they developed a model of serial auto-correlations of the returns. According to them, the fact that returns are auto-correlated tends to show that these returns are linked with a liquidity risk. Indeed, the intuition is that there cannot be auto-correlation in the returns if the market is perfectly efficient. However, if the market is not efficient, it means that there are some frictions in the market. We can think about transaction costs. But one of the most usual frictions on the market is the illiquidity. However, this explanation is relatively vague, this is why there is a second reason to explain this relationship. Indeed, the more the assets are illiquid, the more the managers have a certain leeway for the reporting of their quotation.

With this leeway, they tend to smooth the returns of the fund by reporting smaller ones than reality when these returns are positive. Consequently, they have the ability to make up for potential future losses. They thus conclude that the strength of auto-correlation is linked with this liquidity exposure. Getmansky et al. insists on the fact that these two explanations are complementary in terms of quantitative impact. The fact that returns are smoothed reduces the volatility of the return patterns but at the same times increases the auto-correlation. This is intuitively easy to understand: if the return observed now is partially composed of the past returns, then there is a link between this return and the return in the past. The model they provide illustrates this link. In addition, they stress the fact that in absence of market frictions, these returns have few chances to be smoothed. Therefore, the serial auto-correlation is a perfect measure of illiquidity, both theoretically and for its practical use. A formal description of the model is available in the Appendix 7.2.

From the work of Getmansky et al. (2004), we can identify three explanatory variables for the model we could use: the smoothing index, the aggregate measure of illiquidity, derived from the working paper of Chan et al. (2005), and simply the auto-correlation coefficient.

The smoothing index The smoothing index is defined this way:

$$\xi = \sum_{j=0}^{k} \theta_j^2.$$

This is actually the value of c_σ^2 in the model described in Appendix 7.2 that represents the explanation of the gap between the volatility of the true returns and the observed ones. This gap is even greater when the value of ξ is lower. This value is the concentration of the parameters θ_j, also known as the Herfindahl index. It means that the greater the smoothing effect is, the lower this value will be. If $\xi = 1$, it means that $R_t^o = R_t$ with $\theta_0 = 1$ and thus that there is no smoothing effect.

Aggregate measure of illiquidity Another measure that is introduced is the aggregate measure of illiquidity, measuring the illiquidity for the whole sector. It is defined as

$$\rho_t^* = \sum_{i=1}^{N_t} \omega_{it} \rho_{1t,i}$$

where

$$\omega_{it} = \frac{AUM_{it}}{\sum_{j=1}^{N_t} AUM_{jt}}.$$

N_t is the number of funds and AUM_{jt} is the asset under management for the fund j at time t. $\rho_{1t,i}$ is the first order auto correlation of the fund i using a rolling window on the past returns.

Auto-correlation Finally, a simple measure would be the auto-correlation coefficients. It has the advantage to be very simple to implement and test. On the other side, a choice has to be made regarding the lags of auto-correlation and, if several lags are chosen and thus several measures are derived, a method has to be found to aggregate those in one single measure. Nevertheless, Bussière et al. (2014) worked in the same kind of context as ours as they needed a proxy for the illiquidity exposure of hedge funds to use as an explanatory variable. Inside their model, they investigate its relationship towards the commonality behaviour of hedge funds. To do so, the proxy they use is the simple auto-correlation of lag 1, which would be positively correlated with the liquidity risk of a given hedge fund.

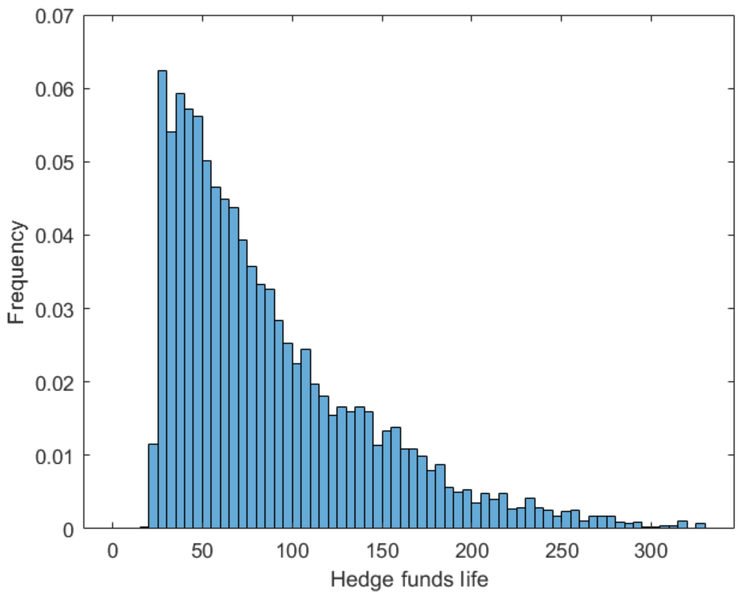

Figure 5: Frequency distribution of the individual hedge funds' lifespan

From this overall picture, it has been decided to use the simple lag 1 auto-correlation to measure this proxy of liquidity exposure that will be used as an explanatory variable. Indeed, the smoothing index involves the implementation of a relatively complex model for an explanatory variable inside a model that is also complex, making the underlying uncertainty too important to handle. On the other side, the aggregate measure of illiquidity uses the very same auto-correlation of lag 1 but with a rolling window. Moreover, the calculation of any proxy at the individual level of hedge funds would require to get rid of the hedge funds with small lifespan, involving a selection bias. The auto-correlation coefficient of lag 1, without the use of rolling window, allows to optimally mitigate this problem as the other solutions would require more reporting data per individual fund and therefore the suppression of more individual hedge funds from the dataset. For this perspective, only hedge funds with a minimum lifespan of 48 months has been selected, involving a minimum of 48 data points per hedge fund and a reduction of the data of 11% from 897,084 to 795,303 observations. Figure 5 gives a better idea of the distribution of the hedge funds according to their lifespan. We clearly see that the distribution is exponentially decreasing, involving that even a selection with a slightly higher threshold would have a big impact on the final sample size.

5 Empirical results and analysis

5.1 LASSO-GP Regression

5.1.1 Penalization approach

The first thing to do is to select the threshold above which one can use the GP regression: the scale parameter (u). In this framework, it has been decided to look at the stability of the different estimated parameters for a simple fit of the GP distribution on the data according the the different possible values of the scale parameter (ξ). On this basis, the threshold is calculated using a quantile of 95% on the negative returns.

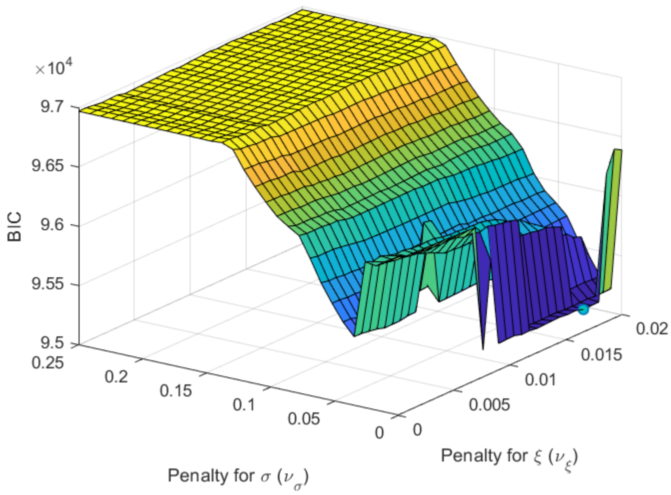

Figure 6: BIC depending on the possible combinations of penalties (ν_σ and ν_ξ)

As explained in section 3.3.1, a grid of the possible values of the penalties must be created and evaluated (we have to maximize $\mathcal{L}_{pen}$ as defined by equation (11) but we need first to approximate the optimal values of $\nu = (\nu_\sigma, \nu_\gamma)$). To do so, the grid is created by defining lower bounds and upper bounds for the penalties before setting the size of the grid. These lower and upper bounds are defined such that they respectively include or exclude all the variables, in order to have the two extreme cases for which it is not useful to go beyond. It is important to bear in mind that the computational effort required for this step can dramatically increase as the size of the grid increases (for instance, a grid of size 5x5 takes approximately 25 minutes for Matlab to process). The selection of the final values of the penalties is represented on the

Figure 6 using a grid of size 30x30, where the minimum BIC penalty combination is pointed. Generally speaking, it is often better to have an optimal combination that is not on the edges represented by the bounds of the grid. Though the figure clearly shows that we are in this case for ν_σ, the fact that the lower bound involves the inclusion of all the variables implies that we have no other choices. Moreover, the fact that σ tends to be unpenalized while ξ tends to be extremely penalized is coherent with the expectations: generally, the shape parameter is more constrained than the scale parameter.

5.1.2 Results of the regression and analysis

With these penalties, we can numerically maximize $\mathcal{L}_{pen}$ from equation (11) through the parameters and obtain the results described on Table 2.

Table 2: Results of the Lasso-GP regression with estimates and p-Values

	Coefficients σ		Coefficients ξ	
	Estimate	p-Value	Estimate	p-Value
constant	1.268	0.000	0.115	0.000
Assets	-0.081	0.000	0.016	0.003
Management Fee	0.043	0.000	-	-
Incentive Fee	0.076	0.000	0.018	0.003
Lockup	0.057	0.000	-	-
Redemptions	-0.023	0.001	-	-
Advance Notice	-0.038	0.000	-	-
CPI	0.060	0.000	-	-
Ind. Prod.	0.013	0.023	-	-
Unemployment Rate	-0.096	0.000	-	-
VIX	0.233	0.000	-	-
MSCI	0.136	0.000	-0.074	0.000
EPU	0.050	0.000	-	-
Negative Lag Returns	0.171	0.000	-0.038	0.000
Auto-correlation 1	0.012	0.055	-	-
Strategy: Equity Hedge (1)	-0.042	0.000	-	-
Strategy: Event driven (1)	-0.024	0.001	-	-
Strategy: Fund of Funds (1)	-0.136	0.000	-0.013	0.005
Strategy: Macro (1)	-0.015	0.124	-	-
High-water mark (yes = 1; no = 0)	-0.039	0.000	0.023	0.000
In emerging markets (yes = 1; no = 0)	0.074	0.000	-0.022	0.000
Leverage (yes = 1; no = 0)	0.032	0.000	-	-
Offshore Vehicle (yes = 1; no = 0)	-0.058	0.000	-	-
LLF	9.484E+04		ν_σ	0.000
BIC	9.519E+04		ν_ξ	0.017

This table gives the estimates α_l^σ and α_l^γ for the l^{th} covariable. Thanks to equation (10), we know that negative impacts on ξ and positive ones on σ would both increase the VaR of a given hedge funds at a given time. Nevertheless, the penalization process requires to scale the explanatory variables, involving that the values of the estimates cannot be interpreted *per se*. However, the signs and, in a relatively cautious way, the differences between the estimates can be analyzed.

Table 3: Impact table on the VaR of Hedge Funds

	VaR impact	
with $\Delta \geq 0$	Through σ	Through ξ
ΔAssets	-	-
ΔMSCI	+	+
ΔNegative lag return	+	+
ΔHigh-water mark	-	-
ΔIn emerging markets	+	+
ΔAC1	+	
ΔManagement fee	+	
ΔVIX	+	
ΔLeverage	-	
ΔAdvance Notice	-	
ΔLockup	+	

Therefore, a positive coefficient for a given variable in the summary table of σ will increase, all things staying unchanged, the VaR but also the Pure Tail Risk (PTR) of a given hedge fund as the observed value of the explanatory variable increases. On the contrary, the negative coefficient estimates for ξ will tend to increase the very same VaR.

Table 3 provides a more synthetic view of the impacts' signs on the VaR involved by the variables that will be discussed. There are variables for which the net impact cannot be identified, such as the incentive fee, for which we have a positive correlation with the VaR through σ but a negative one through ξ. These cases will not be covered as, when it comes to the net impact, the relevant moment to discuss it will be for the computation of the systemic risk. The analysis will thus focus on some of the variables but it is first important to separate the interpretation of these respective parameters into two parts:

- The interpretation of external variables, that will therefore apply the same way to every hedge funds and will therefore be depending only on time. The VIX or the MSCI returns are perfect examples.
- The interpretation of intrinsic variables, meaning what can explain discrepancies among the hedge funds and not only across time. Note also that we can identify hybrid values, those that are actually external to hedge funds but are also different depending on their characteristics. We can for example think about the unemployment rate that will be the one from the referenced country of registration of a given hedge fund.

Regarding the external variables, we first see that the VIX and the MSCI are significant variables in its determination. The VIX can be seen as an indicator of uncertainty in the financial markets. This indicator finds peaks during crises and financial or economic shocks and will affect every hedge funds the same way, since for a given time, the VIX is the same for every hedge funds. It is therefore a good explanation also for the global behaviour of the pure tail risk that is the object of the next section. The Economic Policy Uncertainty (EPU) provides also the same kind of information and is also significant. Here, it shows nothing but what we could expect: the higher the uncertainty on the market, the higher will be the 1% chances losses. For the MSCI, the coefficients tell us that a higher return of the MSCI would increase the VaR of the hedge funds, through both σ and ξ. This conclusion leads us to think that globally the hedge funds tail exposure is short on the global returns, which is also correlated with a narrower equity index, such as the S&P500. We see here a real link with the option-like behaviour developed by Fung and Hsieh (2004) but applied on the lower tail of the return distributions. More particularly, in terms of tail exposures dynamic, this could drive also the potential anticipation effect stressed by Agarwal et al. (2017) that will be discussed later during the analysis of the PTR.

Regarding intrinsic variables, we can first observe that the lagged returns take an important part in the determination of σ as well as ξ, though the penalty for ξ is high. It can be interesting to catch a GARCH-like effect but it is actually not the only reason why such variable could be important. In addition, the GARCH model assumes an auto-correlation in the square of returns while here we look at simple lagged values. Therefore, the importance of such variable could not be fully explained by this effect. Indeed, the fact that current extreme returns can be explained by past returns is intuitively linked with a kind of auto-correlation in the simple returns (all of them, so not only the extreme ones). In this analysis, the fact that lagged values of returns are part of the explanation of the tail exposure could be linked with such liquidity risk as suggested by Getmansky et al. (2004). Moreover, the values of the estimated coefficients for σ and ξ, respectively positive and negative, show that an extreme value in the past increase the chances of extreme values now. The interpretation of this coefficient is therefore very complex but shows that there could be something hidden behind. However, by already catching a hidden liquidity problem, this variable could repeat part information contained in the lag 1 auto-correlation, involving its lower significant in the analysis. However, the penalization procedure kept it for the explanation of σ, which shows that it has still something to tell. Other factors could also be interpreted as dependent to liquidity exposure, potentially creating some endogeneity problems too. Leverage is a good example. The use of leverage is also subject to liquidity risk and is important for the determination of σ and, with its positive value, tends to increase the VaR. Related to that, the lockup period shows interesting results, as it indicates the time during which a new investor cannot take back his or her investment, involving also a greater leeway for liquidity management purpose. However, the fact that a longer lockup period tends to increase the VaR of the funds could be understood as, on the contrary, an indicator that the hedge fund is subject to a bigger liquidity exposure, involving that it needs more time during which the initial investment is locked up. On the opposite side but yet linked with liquidity, we find a negative relationship between the VaR and the advance notice required by hedge funds: a greater value gives more flexibility to funds for liquidity management. It thus involves that, if we imagine that auto-correlation and lagged returns let us suppose that liquidity risk increases the tail risk of hedge funds, then one possible interpretation could be that the advance notice is a good liquidity risk reducer while the lockup period is not. Moreover, the use of high-water mark tends to decrease the extreme exposure. It is a process of performance compensation to protect the investor against situation where he or she has to pay fees while performance is poor, giving incentive to managers to be protected against extreme losses. On the opposite, management fee tends to positively affect tail exposures, involving that managers would potentially tend to take more tail risk when the reward is high. Both variables tend to converge to same interpretation: are hedge funds managers tempted to take more risk when the possible reward is higher? Even if such interpretation can be too fastly drawn, the fact that high-water mark is important in the dynamic of the tail exposure of hedge funds is consistent with the work of Aragon and Nanda (2012). Agarwal et al. (2017) also had the same kind of conclusion and support the same implication from the use leverage in hedge funds.

5.1.3 Approach by strategy

In addition, we see from Table 2 that all the main strategies are selected by the model to explain σ. As a result, it appears useful to fit the very same model on datasets separated by strategies. Through this analysis, we see that Relative Value hedge funds tend to be the most exposed to tail risks (as all the estimates coefficient of the dummy variables are negative), but that does not provide an explanation underlying the difference between this strategy and the other ones, in case there is any. Figures 7 and 8 give part of the answer as the first one gives the VaR per strategy with the model fitted as provided in Table 2, while the second figure gives the same information by fitting the model on all the observations characterized by the same strategy, the hedge funds strategy being then not only a mere variable but the criterion to create 5 different samples.

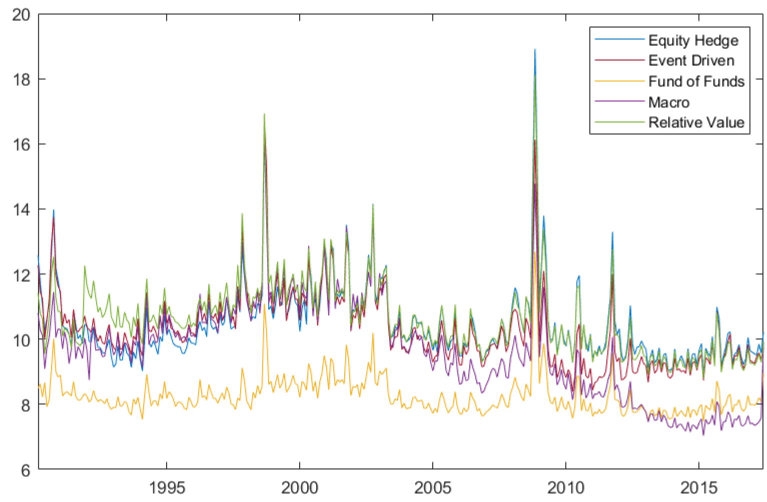

Figure 7: Inferred VaR for the hedge funds depending on the strategies, based on the results of the models shown in the present section

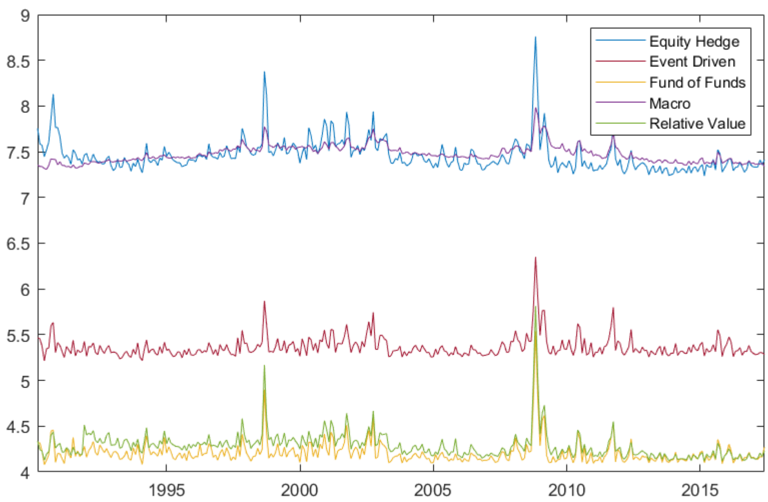

Figure 8: Inferred VaR for the hedge funds depending on the strategies

Figure 8 clearly shows a nearly perfect parallel shift between the strategies. The potential explanation is threefold: (i) there is no clear explanation of what could explain the differences between the strategies; (ii) there is an explanation but it is beyond the scope of the scope of information the HFR database could provide; or (iii) the explanation is possibly available within the HFR data but the sample sizes of the 5 individual datasets (by strategy) are not sufficient to identify it. Except for the first explanation, the only way to solve this could be to extend the data (with other sources), which is not possible for this work. Nevertheless, if we look at the difference of VaR regarding the Relative Value hedge funds, the difference between the two approaches is significant, which could be involved by the poor fit of the model due to the small sample size. At the end, no matter the explanation, the optimal choice for the remaining process is to use the model of the whole dataset, as we have this parallel shift involving that the regression of the whole dataset catches most of the information. This parallel shift is mostly caught by the difference between the constants but also the selection of variables the different hedge funds have in common, extrinsic variables such as the VIX of the MSCI. However, the selection of the variables, as displayed in the Appendix 7.3, can be of interest. We see real differences regarding the explanation through the lag 1 auto-correlation as it is not selected for the Event Driven hedge funds or not significant for Relative Value funds, on the contrary of the other ones. Moreover, the impact of such auto-correlation is even negatively related with the VaR for Funds of Funds. There remains thus Equity Hedge hedge funds and Macro hedge funds. We can notice for the first one the importance of other liquidity related factors, such as the relatively high estimate for negative lag returns. Moreover, Macro funds explanation through sigma is strongly penalized compared to other funds, but the auto-correlation is still selected.

5.1.4 Inference

Thanks to this method, it is possible to have a view of the tail exposure across time on the whole hedge funds industry, either by looking only on over-the-threshold data or by inferring σ and ξ and thus the VaR for all individual hedge funds. To do so, we can gather all the VaR that are calculated at a given time, either over-the-threshold or inferred, and compute the key metrics of this distribution of VaR. That is the insight Figure 9 provides with a calculation of the mean of the hedge funds VaR at a given time, weighted by the AUM. The same approach is possible by using the median and high and low quantiles, allowing for a representation similar to a time-varying boxplot.

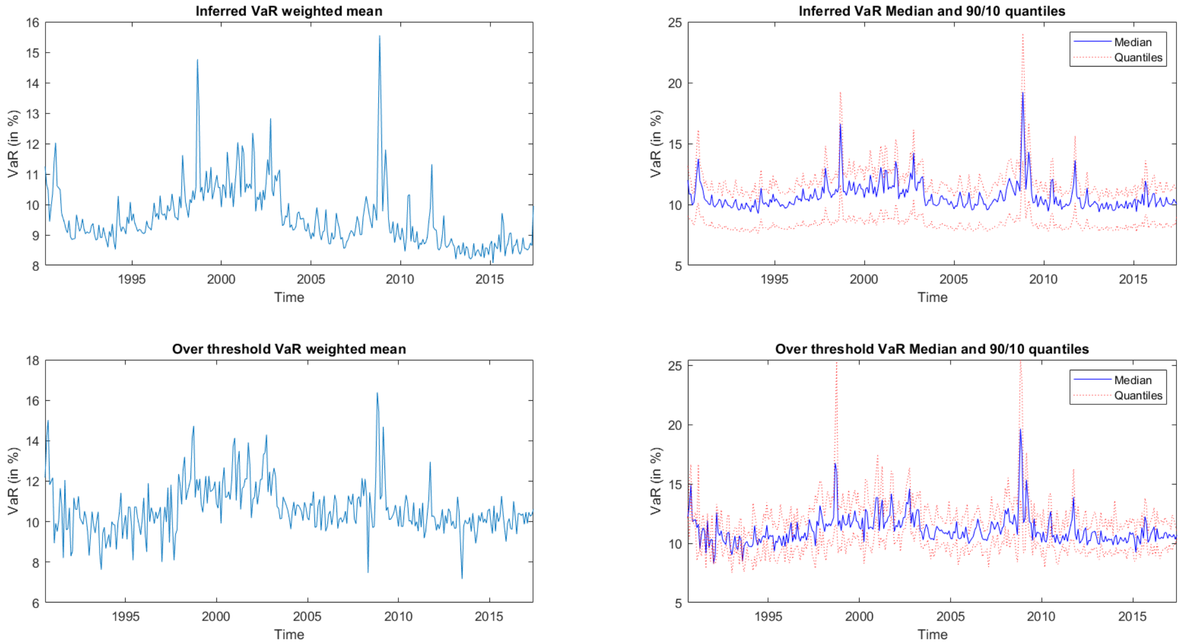

Figure 9: Inferred and over-threshold VaR mean weighted by AUM (left) and inferred and over-threshold VaR median and quantiles (right)

Overall, the graphs show that peaks happen during periods of stress: 1998 with the LTCM crash and the 2008 crisis, as it was pointed to be important crises in terms of contagion by Billio et al. (2010). Between these two specific moments, the overall VaR of the industry would tend to decrease. In comparison, we can see a lower inferred VaR weighted mean in aggregate than the ones for over the threshold data, which is quite logical. However, this is less clear in the case of the median. First of all, the median is globally higher than the weighted mean, involving that the weighted mean could be influenced by extremely low values of VaR. This intuition is also logical as we have observed the negative correlation between the VaR and the size of the funds, involving a weighted VaR pushed down more easily. The median seems also more stable overall, making for example the LTCM accident in 1998 less easy to identify, as it is also the case for the over-the-threshold VaR. Actually, the fact that the inferred VaR is able to catch particularly well this crisis could be due to the VIX, while being in average lower in the periods that followed, involving that other factors would have maintained particular hedge funds exposed to some specific tail risks, typically smaller hedge funds or hedge funds exposed to more liquidity exposures.

5.1.5 Scenario and focus

The model enables also to look into details and compare particular hedge funds across time but also to investigate the VaR distribution of the industry at a given time and comparing it with the same distribution at another date. This is the insight that this section attempts to provide and explain.

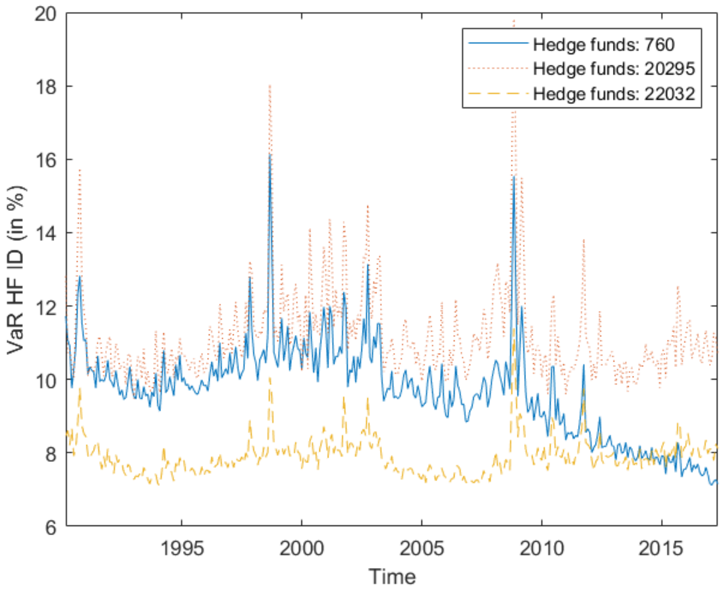

Figure 10: Evolution of the VaR for three different hedge funds

Figure 10 illustrates the VaR evolution of three different hedge funds. These were selected on a random basis among the hedge funds for which the lifetime was maximal, but also by ensuring an intuitive interest in the explanation of their differences. Table 4 provides the exact data of these funds for 2 dates, one where hedge funds 760 and 20295 are close in terms of VaR and another one where hedge fund 760 reach the level of hedge fund 22032. The similar level of tail exposure for hedge funds 20295 and 760 at the beginning of the period, which are higher in comparison to the third fund, can be attributed to country specific characteristics. Indeed, the two hedge funds are registered in the same country. Moreover, the fact that hedge fund 20295 VaR tends to be higher than the 760 one is mostly explained by the lower AUM, as the size is negatively linked with the VaR, but also the management fees that are set to zero for hedge fund 760 (see table 4). However, over time, we observe that the hedge fund 760 VaR tends to decrease to reach the level of the lowest VaR hedge fund. The explanation should therefore come from an intrinsic variable that is fluctuating over time. The most plausible explanation is the size of the hedge fund. Indeed, we observe that the size of this fund is multiplied by around 66, involving a lower VaR.

Table 4: Focus on the data of 3 hedge funds for two particular dates

Date	31-03-95	31-03-95	31-03-95	30-06-14	30-06-14	30-06-14
Assets	92.000	467.506	32.600	6158.000	151.992	2.390
ID	760	22032	20295	760	22032	20295
Management Fee	0	1	1	0	1	1
Incentive Fee	20	0	20	20	0	20
Lockup	0	0	365	0	0	365
Redemptions	90	90	90	90	90	90
Advance Notice	90	95	30	90	95	30
CPI	0.004	0.003	0.004	0.003	0.000	0.003
Ind. Prod.	-0.002	0.000	-0.002	0.004	0.001	0.004
Unemployment Rate	5.4	8.7	5.4	6.3	6.3	6.3
VIX	13.370	13.370	13.370	11.570	11.570	11.570
MSCI	-4.642	-4.642	-4.642	-1.646	-1.646	-1.646
Negative Lag Returns	-1.950	-0.459	-3.130	-1.370	-1.992	3.980
Auto-correlation 1	0.419	0.248	0.119	0.419	0.248	0.119
Strategy: Equity Hedge (1)	0	0	1	0	0	1
Strategy: Event driven (1)	0	0	0	0	0	0
Strategy: Fund of Funds (1)	0	1	0	0	1	0
Strategy: Macro (1)	0	0	0	0	0	0
Strategy: Relative Value (1)	1	0	0	1	0	0
High-water mark (yes = 1; no = 0)	1	0	1	1	0	1
In emerging markets (yes = 1; no = 0)	0	0	0	0	0	0
Leverage (yes = 1; no = 0)	0	1	0	0	1	0
Offshore Vehicle (yes = 1; no = 0)	0	1	0	0	1	0
Inferred VaR	9.872	7.502	10.401	7.805	7.622	10.597

Additionally, it is also possible to look at the distribution of all the hedge funds VaR at a given time, as displayed in Figure 11 for a period of relative stress (September 2011) and a more peaceful time (October 2006). The boxplot clearly shows that the VaR has a wider range of fluctuation in 2006. Since more stress implies more volatility on markets, this volatility can also be found in the tail distribution. However, what the box-plot does not show is that we have systematically two peaks in the distribution, no matter the period. By looking at the characteristics of the observations concentrated in the 2 peaks of the 2 distributions, we can identify a common feature: the main strategy. When we look at the observations concentration in the VaR intervals $[13, 15]$ and $[9.5, 11]$ for September 2011 and $[9.4, 10.2]$ and $[7.4, 8.2]$ for October 2006, we find a high density of Equity Hedge hedge funds in the high VaR interval (47% for September 2011 and 56% for October 2006) while being almost absent in the lower peak (representing only 1% of the observations). Even more noticeable, with the very same comparison, the concentration of Funds of Funds reaches a peak of 96% and 99% in the lower intervals respectively for September 2011 and October 2006. This stresses again the discrepancies that exist between the hedge funds styles.

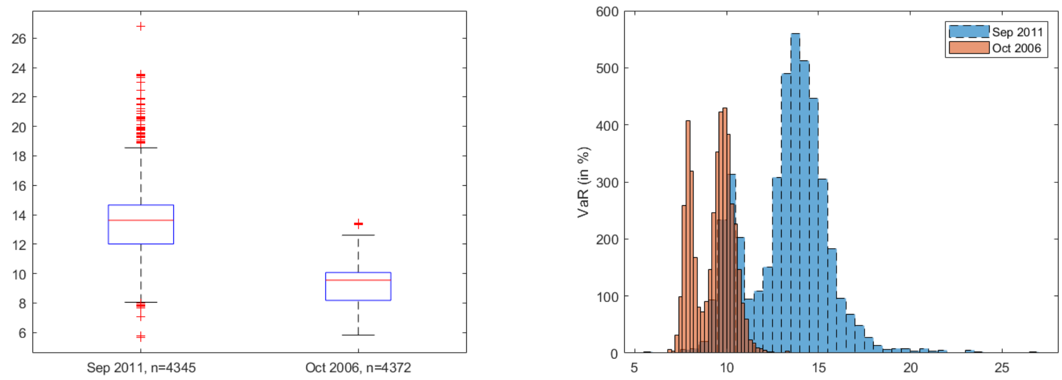

Figure 11: for two different dates, Box plot (left) and distribution histogram (right) of the hedge funds VaR

5.2 Pure Tail Risk

For a reminder, the Pure Tail Risk (PTR) is the ratio between the VaR of one hedge fund at a particular time and the VaR of the equity index at the same time ($\frac{VaR_t^i(p)}{VaR_t^s(p)}$). To retrieve the PTR, the VaR of the equity index is computed thanks to the EVT-GARCH that has been described earlier. Figure 12 illustrates the VaR resulting from the computation.

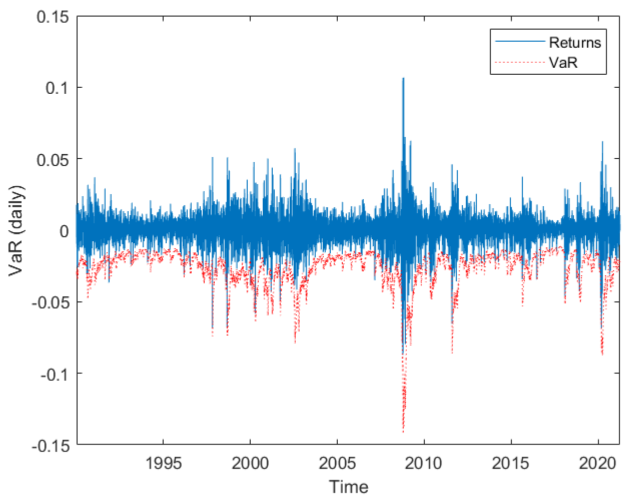

Figure 12: Daily VaR and returns of the S&P500

Several lags for of the GARCH/ARCH degrees have been investigated and are described in the

Appendix 7.4. The GARCH(2,2) has been chosen as all the coefficients were estimated to be significant. We then obtain the monthly VaR by aggregating the daily estimates of the same month. The choice of GARCH might result in a slight overfit of the data as the developers of this method judged that the GARCH(1,1) could be sufficient. However, the objective of this analysis is to provide the most accurate measurement and explanation of the past and not a VaR forecast for risk management purposes. This involves that neither the avoidance of an over-fitted model, nor a back-testing of the VaR is the priority. However, avoiding an overfitted model will be the objective when we will have to treat the tail dependence problem thanks to the regression tree or the Probit model.

In the very same fashion as for the VaR of the hedge funds, we can illustrate the evolution of the PTR thanks to the inferred VaR of hedge funds but also focus only on the over threshold observations on which the model has been fitted, as this is illustrated by Figure 13. As we stated previously, the PTR can be seen as the relative scale of the extreme exposure towards the equity market. Therefore, when it is bigger than 1, one possible interpretation is that extreme shocks on the market will have a bigger effect on hedge funds than it could have on the equity market, increasing also the systemic risk coefficient (see equation (2)). Here we will focus on the weighted mean of this PTR by month, since, as we also mentioned before, it smooths the variability due to intrinsic hedge funds characteristics and can thus be a faithful representation of the whole population at time t or, in other words, the whole hedge funds sector. First, the value fluctuates around 1 but tends to exceed this value, which implies an increase in the systemic risk coefficient of hedge funds for a given value of Systemic Linkage. What is interesting is that there seems to be a sort of disynchronization between the PTR changes and the shocks. For example, the 2008 crisis took place more particularly in September 2008 while we see a huge drop of the tail exposure of hedge funds relative to the S&P500 which is actually measured in August of the very same year. Again, many interpretations are possible but this is also consistent with the work of Agarwal et al. (2017) in which they tend to prove that hedge funds were able to anticipate such movement and thus reduce their tail exposure in times of stress, such as the 2008 crisis. On the contrary, before the crisis, the sector tends to have a higher tail exposure than the equity market. The study of Bussière et al. (2014) showed that the commonality of hedge funds returns increased significantly between 2003 and 2006, but also found that it came with a high liquidity risk, which could be a driver, as we said before, for the tail exposure of hedge funds. The fact that this relative tail exposure is so high during the exact same period involves therefore a real threat for the financial stability as it couples commonality and high tail exposure.

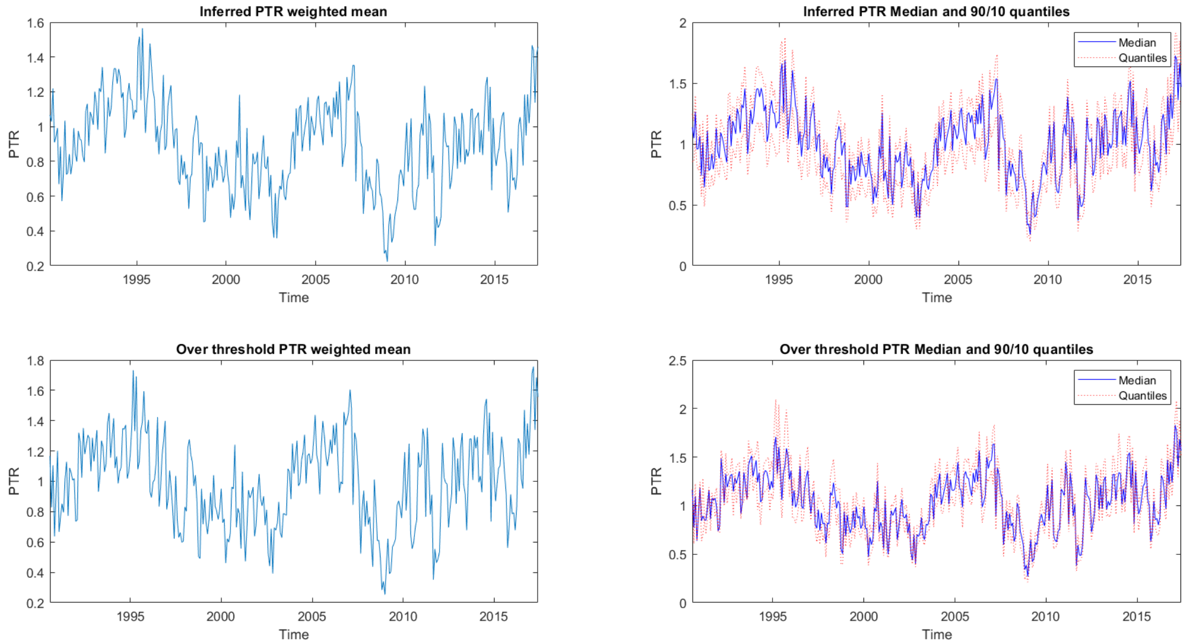

Figure 13: Inferred and over-threshold PTR mean weighted by AUM (left) and inferred and over-threshold PTR median and quantiles (right)

Looking at the evolution of the mean PTR weighted by the AUM can provide an interesting overview of the tendency of the whole hedge fund sector but the question is therefore: is it really representative? With respect to the strategies, it tends to be the case though there is a real differentiation in terms of hedge funds VaR, as it was discussed earlier. Therefore, it does not alter the above interpretation of the PTR, as we observe the same trend as the Figure 14. However, the Funds of Funds tend to be, especially in the beginning, really at a lower value compared to other strategies, while the end of the time frame gives a less clear picture. Nevertheless, whatever the strategy, even if there is a parallel shift in VaR, it confirms the fact that whole industry is subject to this anticipation effect.

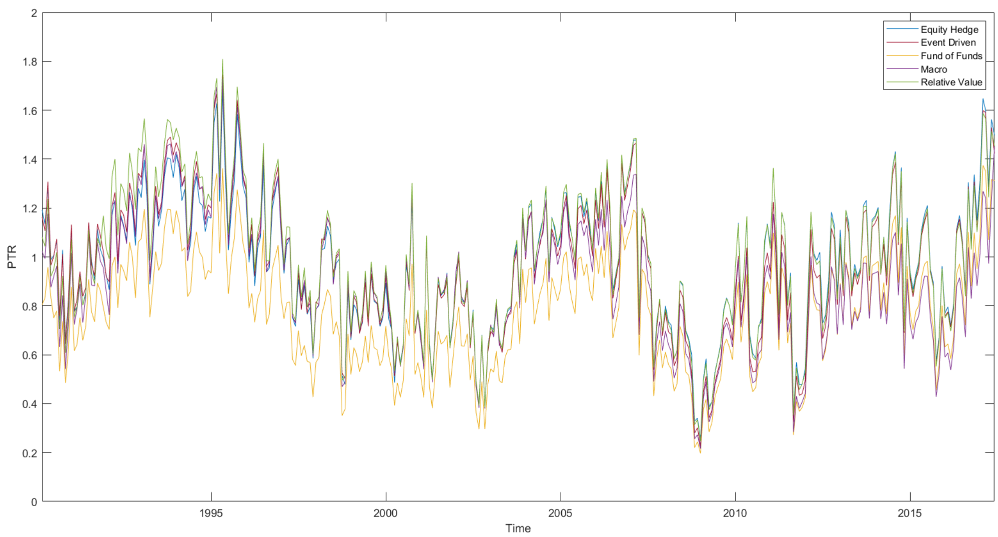

Figure 14: PTR evolution by strategy

5.3 Systemic Linkage

Now that we have the PTR, the right part of equation (2), the left part remains: the Systemic Linkage (SL), or to refer to the very same equation: $\tau_i(p)^{1/\zeta_t^s}$. For a reminder, we will assume that $\tau_i(p) = \lambda_{L,i}$, being the lower tail dependence estimated through a copula distribution.

5.3.1 Tail dependence for individual hedge funds

The Systemic Linkage represents the main challenge as it involves the calculation of the lower tail dependence of an hedge fund with respect to the equity index. The first approach was to focus the analysis only on hedge funds with a sufficient lifetime (more than 100 months) and try to compute it in a non-parametric fashion thanks to the methodology of Embrechts et al. (1999). However, the results were still suffering from the lack of data, mainly induced by the low number of S&P500 data points, as these are monthly returns. The main problem was not only the short lifetime of the individual hedge funds but also, and more importantly, the frequency of reporting. Therefore, increasing the lifetime threshold for hedge funds selection would not resolve the problem and would even create more selection bias. The use of copula instead, for a parametric approximation of the tail dependence, would not only solve the problem but would also allow us to keep the current dataset (a lifetime threshold of 48 months as explained in section 4.2.2).

Given what the Matlab function *copulafit* provides, three copula distributions were tested: Student, Clayton and Gumbel copulae. The third one was rapidly excluded due to its particular poor performance. The two remaining were fitted on individual hedge funds. In terms of goodness of fit, if we look at the average AIC and BIC of the individual fits, the AIC gives a better fit to the Clayton copula while the BIC points the Student copula as the best choice. However, the goodness of fit captures the whole density while the main matter here is the distribution in the lower tail.

	Average AIC	Average BIC
Clayton	-19.466	-20.773
Student	-19.199	-21.913

Table 5: AIC and BIC for Student and Clayton copula

To estimate the tail dependence from the fit of the copula distributions, asymptotic estimations exist. The Student copula is characterized by two parameters: the correlation coefficient (ρ) and the degree of freedom (ν). The relation between the tail dependence and these parameters is

$$\lim_{p\to 0} \tau_i^t(p) \approx \lambda_{L,i}^t = 2 - 2t_{\nu+1}\sqrt{\frac{(\nu+1)(1-\rho)}{1+\rho}}. \tag{18}$$

On the opposite, the Clayton copula is described by only one parameter that we will denote here θ. There is also an asymptotic estimation for this which is given by

$$\lim_{p\to 0} \tau_i^t(p) \approx \lambda_{L,i}^{Clayton} = 2^{-1/\theta}. \tag{19}$$

When we sort the tail dependence in increasing order, we obtain the results displayed on Figure 15. It clearly indicates a higher number of zero values in the case of the Student copula, which might mean a certain inability from the Student copula to catch the join tail distribution of the two returns, while the Clayton copula suffers less from this problem. Moreover, when we investigate the potential relationship between the length of life of the hedge funds and their tail dependence, we find a significant positive relationship, even stronger in the case of the Student copula. We can thus attribute at least partially the fact that tail dependence values fall at zero to the short time of reporting of concerned hedge funds. If this assumption is made, it means that some tail dependence values are set to zero while they should not. Moreover, and this aspect is even more important, the global model of systemic risk we estimate assumes that there is a tail dependence. This assumption can actually be confirmed by the fact that the Clayton copula generates a large majority of non-zero tail dependence. Therefore, if we simply use these zero cases while there should be a tail dependence, we generate zero systemic risk coefficients while, actually, they should not be set to zero. For all these reasons, it is important to find a way to get rid off these zero values. However, these are still estimates of the true values and should, at least partially, represent the truth. The re-estimation of these hedge funds tail dependence is thus expected to generate, for a large part, low values of tail dependence.

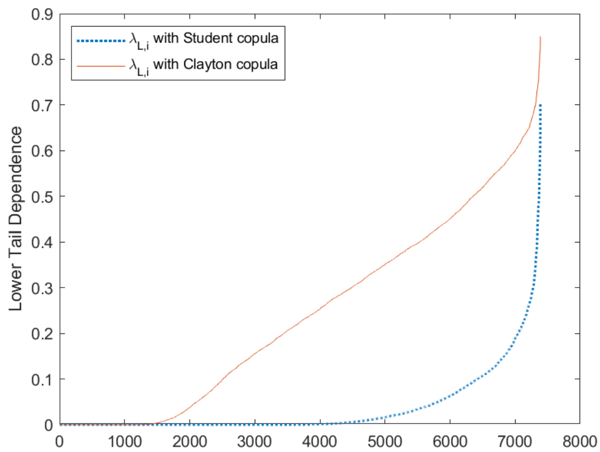

Figure 15: Hedge funds lower tail dependence computed on an individual level based on Student and Clayton copula

5.3.2 Regression tree approach

To solve this issue, groups of hedge funds are created that have common characteristics, identified to be relevant for the determination of the tail dependence. Once we know these characteristics, we are able to assign each hedge fund to one and only one group. From the aggregation of the data in each group, it is then possible to fit the Clayton copula again and retrieve the tail dependence of the group as a whole. Finally, for hedge funds with a zero tail dependence, the group to which they individually belong are identified and their tail dependence becomes the one of their respective group. There are thus two main assumptions behind this procedure:

- The zero tail dependence is mainly caused by a lack of data involved by the short time of reporting of individual hedge funds

- The actual tail dependence can be explained by funds intrinsic characteristics.

For previously mentioned reasons, this analysis will be performed based on the tail dependence induced by the fit of the Clayton copula. In addition to provide a solution to the aforementioned problem, it can also provide an explanatory power to the systemic linkage estimation.

The remaining question is thus how to build these groups. The approach consists in using a regression tree where the response variable is the individual tail dependence and the dependent variables are the funds intrinsic variables that were in the starting selection of the candidates for the LASSO-GP Regression. The difficulty with regression trees is to find the right depth or number of splits. The common way to evaluate this aspect is to split the data randomly into two separate datasets: a training dataset and a test dataset, the first one being used in order to fit the tree, the second one to evaluate its performance. The fit and the evaluation is done through the minimization of the squared error (MSE). By doing so, as we discussed earlier, we avoid the risk to have an overfitted model. It is here required as we will actually forecast tail dependence for some hedge funds. The depth will thus be defined as the one that would minimize the value of the MSE for the test dataset, which corresponds to a number of split equal to 24, as shown on Figure 16. A table with a description of criteria selected by the tree, called ”split table”, is provided in the Appendix 7.5.

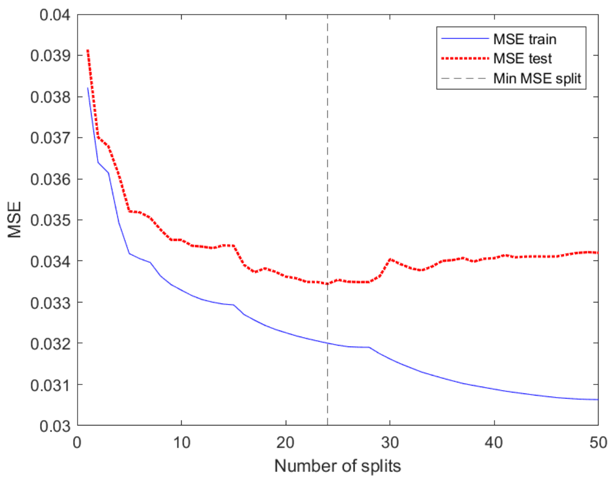

Figure 16: MSE of the regression tree depending on the number of splits

The results of this approach compared with the results we had previously on an individual level are represented on the left side of Figure 19. There are three main observations we can draw. First of all, we have a clear reduction of zero values, which is logically induced by the process. Secondly, there are still hedge funds with a nearly but not exactly zero level of tail dependence (94% of the previously zero cases are now situated at a level below 1%). Finally, the new non-zero values with an increase that makes the tail dependence falls beyond 1% (the 6% remaining) are mainly located in a low range of tail dependence (the highest increase reaches a tail dependence of 41% and represents 0.2% of the old zero cases). From these observations, we can say that the regression tree approach gives consistent results as we do not have transfer from zero to high level of tail dependence though it keeps having hedge funds with very low tail dependence. Though the approach reduces the presence of zero values due to lack of data, we still have a large number of hedge funds with a tail dependence near zero (20%).

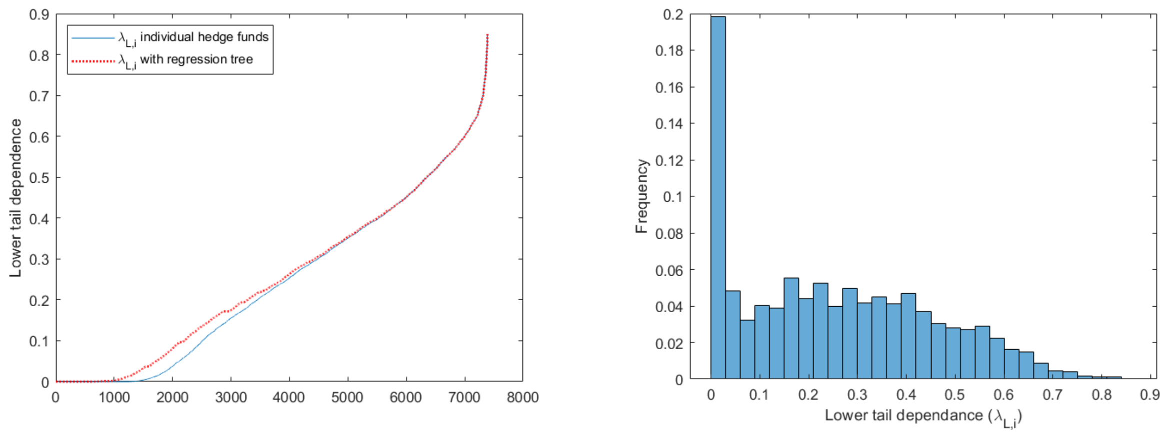

Figure 17: Hedge funds lower tail dependence before and after regression tree procedure (left) and distribution of the lower tail dependence with the regression tree approach (right)

As mentioned during the elaboration of the regression tree procedure, this approach allows to gain some insight in the potential explanations behind the tail dependence. However, the tree counts 24 splits, involving 25 leaves or groups. Giving an intuitive and clear overview thanks to the whole tree could be thus tedious. This is why the first approach is to look at variables that are ”high” in the tree, involving a relative importance. The first and the fifth criteria are related to the strategy, a variable easy to investigate as we have 5 potential values for this variable.

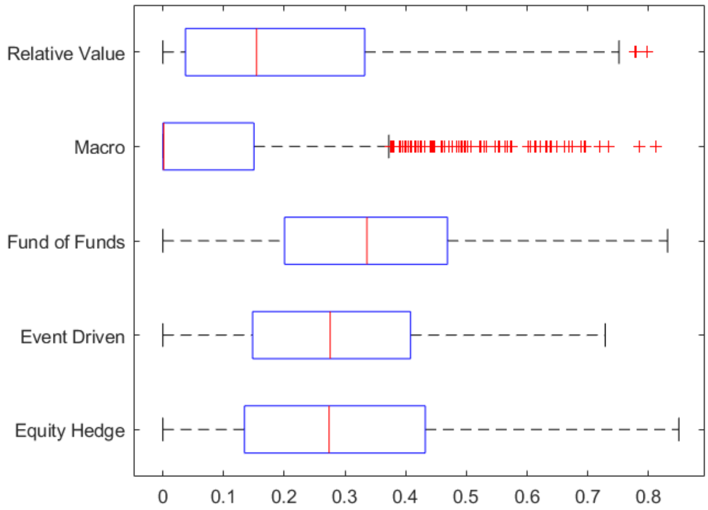

Figure 18: Box plot of the tail dependence by strategy

Figure 18 shows that Relative Value and Macro hedge funds tend to have lower tail dependence than the other strategies. When looking into more details, Macro hedge funds represent 49.35% of the hedge funds with a tail dependence lower than 3% and in the whole population of Macro hedge funds, 61.22% are situated below this threshold. Though there are numerous extreme values with high tail dependence, we can reasonably say that Macro strategy tends to be less linked in the extremes with the system. Nevertheless, this analysis must be interpreted with caution as its rigour could be discussed. In addition, relatively clear results would be hard to apply with this method for other variables, such as numeric (continuous) ones.

Another way to investigate the potential explanation behind tail dependence would be to use another regression tool, easier to interpret. Indeed, now that have generated tail dependence coefficients through the regression tree, we can fit on these values another regression tool. This regression will actually give the very same picture as the one the tree describes but in a way that will be easier to understand and interpret. As the explained variable would be the tail dependence, which is a probability, the most suited approach would be to use a Generalized Linear Model with the Probit function as the link, also known as Probit regression, generally used in decisions forecasts. However, we are not in a context of decision making but rather a simple explanation of a certain quantity. Provided the only objective is the interpretation and not a forecast, we do not have to constraint the response by a link function. As a result, using a simple linear regression could work in such context and would even be easier for the interpretation. We thus estimate both models and the results are displayed on Table 6.

Table 6: Linear and Probit regression results for the interpretation of the tail dependence through regression tree, sample size of 795303 observations

	Probit		Linear	
	Estimate	p-Value	Estimate	p-Value
Constant	0.019	0.000	0.470	0.000
Advance Notice	-0.001	0.000	0.000	0.000
Incentive Fee	-0.016	0.000	-0.005	0.000
In emerging markets (yes = 1; no = 0)	0.067	0.000	0.031	0.000
Leverage (yes = 1; no = 0)	-0.126	0.000	-0.042	0.000
Lifetime	0.000	0.000	0.000	0.000
Strategy: Macro (1)	-0.731	0.000	-0.177	0.000
Strategy: Relative Value (1)	-0.266	0.000	-0.085	0.000
Management Fee	-0.110	0.000	-0.035	0.000
Offshore Vehicle (yes = 1; no = 0)	-0.159	0.000	-0.042	0.000
Redemptions	0.001	0.000	0.000	0.000
Auto-correlation 1	0.067	0.000	0.024	0.000
R^2 adjusted	0.270		0.271	

The results confirm that especially Macro hedge funds are less subject to tail dependence, as it is also the case for Relative Value. Another interesting aspect is the key role of management fee and incentive fee, the first one being represented in four different split criteria in the regression tree and the other one being the second and third criteria chosen by the regression tree after the Macro strategy variable. The fact that the tail dependence tends to decrease should mean that the fund seems to be relatively less linked with the extreme fluctuations, being hedged against it. The higher remuneration of fund's managers could be seen as a proxy of the quality of the manager, even more when it comes to incentive fees as these are remunerations only for good performance. Moreover, we notice that the leverage, if used, tends to decrease the tail dependence. As most of the time, leverage is used by hedge funds to generate market neutral exposures, these results would confirm that it could apply on their tail exposures too. Finally, as the advance notice allows a certain flexibility for the managers in terms of liquidity constraints, the result tends to show that a greater flexibility would reduce the tail dependence. When we think about liquidity spirals essentially due to massive fire-sales in period of crisis and market panic, the fact that investors of hedge funds have to give longer advance notice provides more opportunities for managers to cope with these liquidity problems, as the investors cannot redeem immediately. Oppositely, if there is no advance notice at all, then the investors can take back their money immediately, increasing liquidity exposures for the funds, making it more sensitive to extreme shocks on the market, and thus increasing the dependence to the tail risk of the system. But in terms of liquidity, the auto-correlation lag 1 is also selected, though the coefficient seems high, the tree has selected it relatively lately (criterion n° 21), mitigating its importance.

Nevertheless, a question must be asked: why not using these types of regression instead of the regression tree? In such a context, the objective would have been to provide a forecast and thus the linear model would have been excluded. For the Probit regression, it could be indeed used as we could be able to forecast. Besides, the real insight the regression tree provided was to enable groups creation, on which a new copula could be fitted, benefiting from the aggregation of data. As a result, the question is: it is a sufficient reason to give up the idea of the Probit regression as a solution? The purpose of the next section is to investigate this alternative.

5.3.3 The Probit regression

In the previous section, we grouped individual hedge funds based on their characteristics thanks to the regression tree. In this framework the Probit regression was used only to give another view the tree was already giving. In this section, the objective is not the same as the Probit regression is no more used as a representation of the regression tree but as a substitute to it. Indeed, we give here an actual forecast on the basis of the Probit regression. therefore, using the explanatory variables selected by the LASSO-GP Regression for the PTR, a Probit regression is fitted using only observations with non-zero individual tail dependence. As the tail dependence that was previously computed is constant over the life of one hedge fund, using several reporting observations from one single hedge fund would not be a good idea. To be consistent, we computed the mean of each explanatory variable over the life of each hedge funds. As a result, each observation in the sample is one single hedge funds with its average characteristics. It results in sample size of 7388 observations. The results of the regression is displayed on Table 7.

Table 7: Probit regression results for the estimation of tail dependence, sample size of 7388 observations

	Estimate	p-Value
Constant	0.072	0.402
Advance Notice	-0.001	0.000
Assets	0.000	0.671
CPI	59.948	0.000
EPU	0.012	0.000
High-water mark (yes = 1; no = 0)	0.006	0.788
Incentive Fee	-0.016	0.000
Ind. Prod.	15.364	0.023
In emerging markets (yes = 1; no = 0)	0.020	0.475
Leverage (yes = 1; no = 0)	-0.115	0.000
Lockup	0.000	0.023
Strategy: Equity Hedge (1)	0.317	0.000
Strategy: Event driven (1)	0.276	0.000
Strategy: Fund of Funds (1)	0.257	0.000
Strategy: Macro (1)	-0.256	0.000
Management Fee	-0.132	0.000
MSCI	-0.154	0.000
Offshore Vehicle (yes = 1; no = 0)	-0.113	0.000
Redemptions	0.001	0.000
Unemployment Rate	-0.004	0.515
VIX	-0.094	0.000
Auto-correlation 1	0.207	0.000
R^2 adj.	0.254	

The results confirm the interpretation provided with the regression tree. The impact of the auto-correlation 1 is even stronger. When comparing the two approaches, we see that the Probit model generates even less zero values. It has also the advantage to provide the solution and the explanatory power at the same time. Nevertheless, the insight the regression tree provides that the Probit regression could not is an ordering between the variables, involving for example the nuance underlying the importance of the auto-correlation.

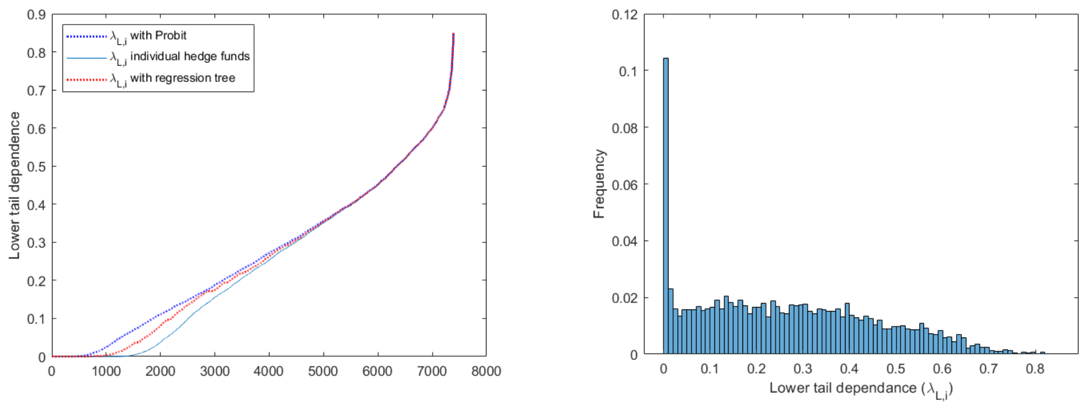

Figure 19: Hedge funds lower tail dependence before treatment, after regression tree and after Probit regression (left) and distribution of the lower tail dependence with the Probit regression approach (right)

Though the impact seems to be important, the two procedures concern the inference of tail dependence for 12% of the individual hedge funds. Whether we choose the regression tree or the Probit model, the final estimation has proven to be robust to the choice of one model or the other. However, since the Probit model (i) does not require two estimations procedure, (ii) gives a relatively appreciable explanatory power (25.4%), (iii) generates even less extremely low tail dependence while still keeping a good part and (iv) allows an easier- and faster-to-interpret approach, this model has been chosen for the final computation of the SL and the systemic risk coefficient.

5.3.4 Systemic Linkage analysis

Once we have the tail dependence, we can compute the Systemic Linkage (SL) thanks to the quantification of the tail index of the system (the S&P500 returns). It is calculated on the monthly returns thanks to a fit of Generalized Pareto Distribution. As showed by Mikosch and Starica (2000), the theoretical shape parameter should be constant even under the assumption of heteroscedasticity of the GARCH model. Therefore, we make the assumption that the tail index of the system, being equal to the inverse of the shape parameter, is relatively constant over time. This assumption allows us not to add another time variation that would not have been easy to interpret. Moreover, an EVT-GARCH approach would require the use of the daily returns, meaning that we would have to find a methodology to aggregate a 'daily' tail index into a 'monthly' one. With the chosen approach, we can directly compute the tail index on the monthly returns, avoiding us to cope with this relatively complex.

As it was also the case earlier, for a given month we are able to compute the mean of the SL of each hedge fund weighted by the AUM. Though the computed tail dependence does not change over the life of each individual hedge fund, there are new hedge funds entering the market, and others that may have died. This creates a time varying fashion to the SL of the whole industry that can be interesting to represent, as Figure 20 does.

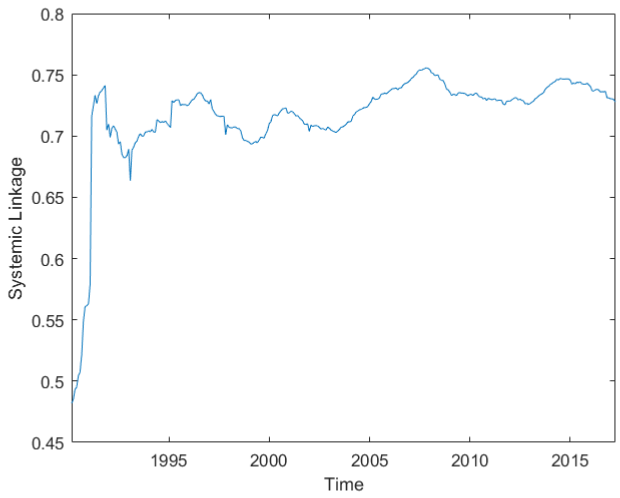

Figure 20: Systemic Linkage mean of the hedge funds industry, weighted by the AUM

Though it looks like the representation we discussed with the PTR and the hedge funds VaR, its interpretation is more tedious for two reasons. First, the time-varying feature of the SL is only created by the birth and the death of different individual hedge funds, while the PTR has this time varying fashion thanks to the regression method. However, that does not mean that this is not representative of the true SL evolution, we can simply see this approach as a rolling window estimation. Secondly, the tail dependence of one hedge fund, being constant over its life, takes into account all the returns it generated. It thus means that when a new fund comes into place, the tail dependence he gets in this analysis takes into account nothing else but its future and, inversely, only the past the day the fund dies. This could explain for example why we have a large decrease after 2008 and not exactly at this time. The time varying feature of this representation is thus difficult to interpret. Yet, the estimation is still able to catch variation across time, allowing us to identify a slight increase starting from 2004.

As it has been shown that the strategy of the hedge funds plays a key role, it is possible that this increase can be driven by one strategy in particular. Though these were not pointed by the earlier analysis, Figure 21 shows that we have indeed a truly lower SL for Macro hedge funds while the Equity Hedge, the Funds of Funds and the Event Driven hedge funds have clearly higher tail dependence. Compared to the analysis of the PTR, the order is not the same, meaning that the final systemic risk coefficient will be influenced in two opposite ways. Consequently, we will have to find which component carries the biggest impact on the final computation.

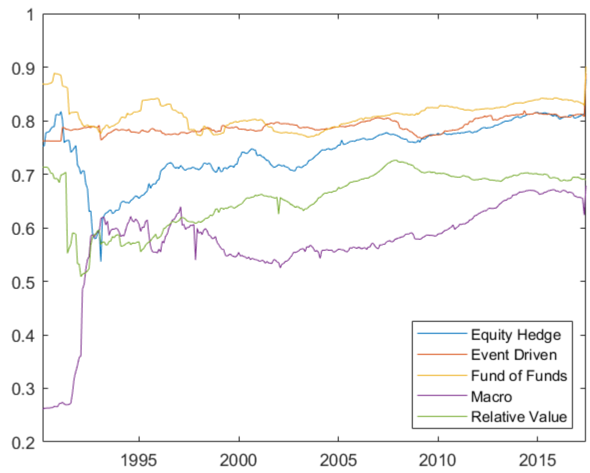

Figure 21: Systemic Linkage mean of the hedge funds industry by strategy, weighted by the AUM

5.4 Systemic risk

As a reminder, the systemic risk coefficient as defined by equation 2 is the multiplication of the PTR and the SL, and therefore also the sum of their logarithms. As we have managed to represent the evolution over time of both measures, the same representation is possible for the systemic risk. Thanks to the time-varying fashion of all the components, we can have a time varying systemic risk coefficient, as illustrated by Figure 22.

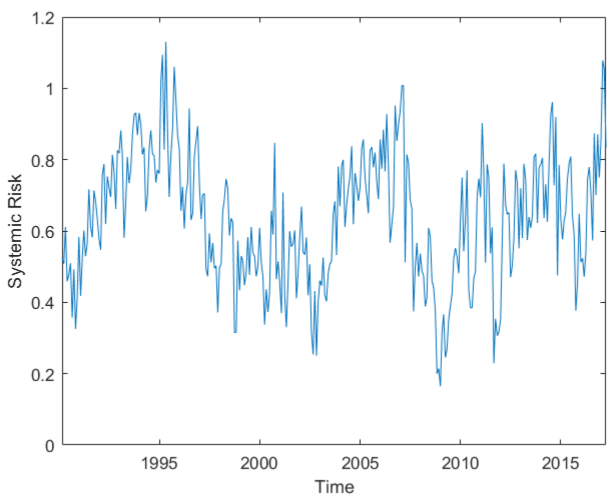

Figure 22: Systemic risk mean of the hedge funds industry weighted by the AUM

The fluctuations can be compared to the ones observed with the PTR, though the SL still should play a role. Figure 23 can give a clearer view of their respective contribution. The increase between 2004 and 2008, discussed during the analysis of the PTR, is mainly driven by this PTR. The same observation applies for the period after the crisis of 2008, during which the systemic risk has been steadily increasing until the end of the time frame. The very principle of the method makes such that the SL fluctuates way less than the PTR over time, involving that the time-varying fashion of the systemic risk is mostly driven by this last element. More importantly, the results of the model show that the level of systemic risk of hedge funds at the end of the time horizon have never reached such high level before. As Bussière et al. (2014) showed that commonality across hedge funds increased a lot during the less volatile period of 2003 - 2006, which corresponds also to high systemic risk, the same could apply to the period that follows the 2008 crisis.

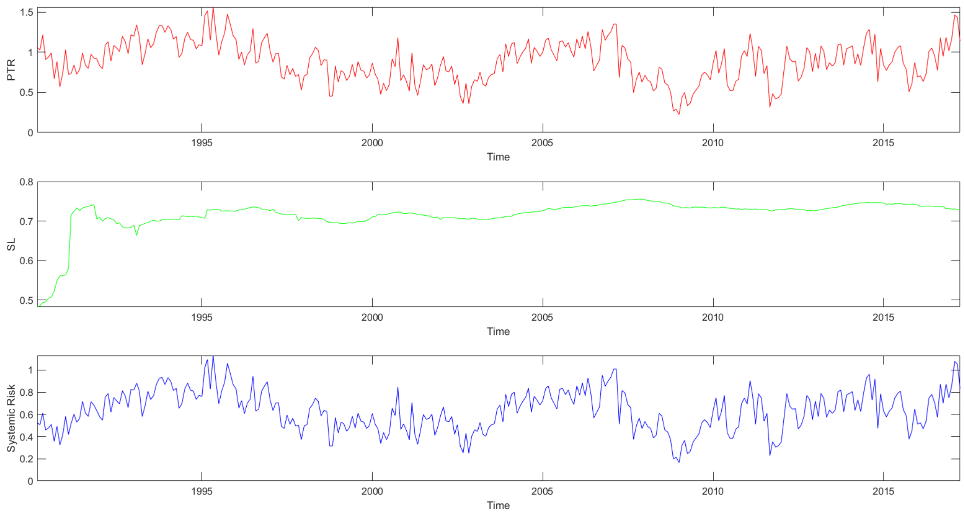

Figure 23: (From top to bottom) PTR, SL and Systemic risk mean weighted by AUM over time

The explanation of the systemic risk as a whole is tedious as it non-linearly combines other measures (VaR of the hedge funds and the equity index, tail dependence, and tail index of the equity index) that were estimated and explained separately. Therefore, two approaches are possible to provide an explanatory power:

- Comparing the explanations of the different measures and see if some of these have a similar impacts on both the SL and the PTR and thus on the systemic risk coefficient.

This consists in the first step.

- In a second step, for variables for which the global impact on the systemic risk cannot be clearly identified in the first step, looking at the computed mean systemic risk over time for hedge funds grouped on the basis of these different characteristics. This approach will give only intuitive results and will not provide any statistical significance, involving that we must be cautious with the implications.

5.4.1 Allignment of the PTR and the SL analysis

When we look at the outcomes of the LASSO-GP regression for the PTR and the Probit model for the SL, we can build a table that could summarize the different impacts signs through the two different models respectively used for the explanation of the PTR and the SL.

Table 8: Identified impacts' signs on the systemic risk coefficient through the two different models (N.Se is for 'Not selected' by the penalization process, N.Si for 'Not significant' in the Probit model)

For $\Delta \geq 0$	PTR		SL	Systemic Risk
	Through σ	Through ξ	Probit model	
ΔAdvance Notice	-	-	-	-
ΔAssets	-	-	N.Si	-
ΔHigh-water mark (yes = 1; no = 0)	-	-	N.Si	-
ΔIncentive Fee	+	-	-	
ΔIn emerging markets (yes = 1; no = 0)	-	+	N.Si	
ΔLeverage (yes = 1; no = 0)	+	N.Se.	-	
ΔLockup	+	N.Se.	+	+
ΔStrategy: Equity Hedge (1)	-	N.Se.	+	
ΔStrategy: Event driven (1)	-	N.Se.	+	
ΔStrategy: Fund of Funds (1)	-	+	+	
ΔStrategy: Macro (1)	-	N.Se.	-	-
ΔManagement Fee	+	N.Se.	-	
ΔOffshore Vehicle (yes = 1; no = 0)	-	N.Se.	-	-
ΔRedemptions	-	N.Se.	+	
ΔAuto-correlation 1	+	N.Se.	+	+

For concision, Table 8 gives an impact summary only for the pure intrinsic variables of hedge funds. The signs reported for the systemic risk coefficients are only those for which we have a clear alignment between the two models.

We have first of all the advance notice, as it was discussed earlier as well as its potential information regarding liquidity management. As a reminder, the variable was selected early by the regression tree, reinforcing its importance. As it has also been discussed with the PTR at the same time, the lockup period is also positively correlated with the systemic risk of hedge funds.

As the lockup period is a period during which the investors cannot redeem their investments, a longer time should allow the managers to have opportunities for a better liquidity management. However, this positive correlation leads to the same conclusion as with the PTR: the lockup period is more a indication of bad liquidity management more than a liquidity risk reducer. For both the incentive fee and the lockup period, the discussion related to the PTR applies also to the SL and thus on the systemic risk.

Regarding the AUM, it also plays an important role but only through the PTR and not the SL. More intuitively, that means that the size of the hedge funds does not condition its potential link in the extremes with the markets but is a good proxy of how good the hedge fund manages the size of this extreme link. We could associate that with an experience effect. This is also linked with the findings of Harri and Brorsen (2004) showing that there is a negative relationship between the returns and the size of the hedge fund. The same thus applies on the tail risk exposure. The same impact is observed for the use of high-water marks. The high-water mark ensures the fact that the manager will not get any extra remuneration that it should not deserve for poor performance. It thus gives an additional incentive for managers to hedge against extreme risks. The analysis tends to show that this aspect plays on the PTR only. It involves that the hedge fund will be, all other things staying equal, exposed to the same extreme linkage with the system, but, again, the size of the exposure involved in this linkage tends to be mitigated.

Regarding the strategy, since the base case is the Relative Value strategy (if all strategy variables are set to zero, then the hedge fund is a Relative Value hedge fund), the impact is with respect to this strategy. Though it seems that Relative Value has the highest PTR, the same does not apply on the SL. However, only the Macro hedge funds clearly shows a systemic risk that is lower. For the other strategies, it does not seem to be possible to clearly identify the overall impacts compared to relative value. Regarding the use of offshore vehicle, we have a negative overall impact. This variable has not been considered yet but tends actually to have a clear impact. Using an offshore vehicle involves that the concerned hedge fund is not subject to its own country tax regulation in order to benefit from other countries taxation. The purpose is also for example to attract non-home-country investors. It is thus logical to witness a certain disconnection with the system. Since USA-hedge funds (representing the majority of the population) using offshore vehicles tend to attract non-USA investors, it is thus possible that the potential fears investors could have during market shocks on the US financial market (such as on the S&P500) could be lowered, involving less panic withdrawals from them.

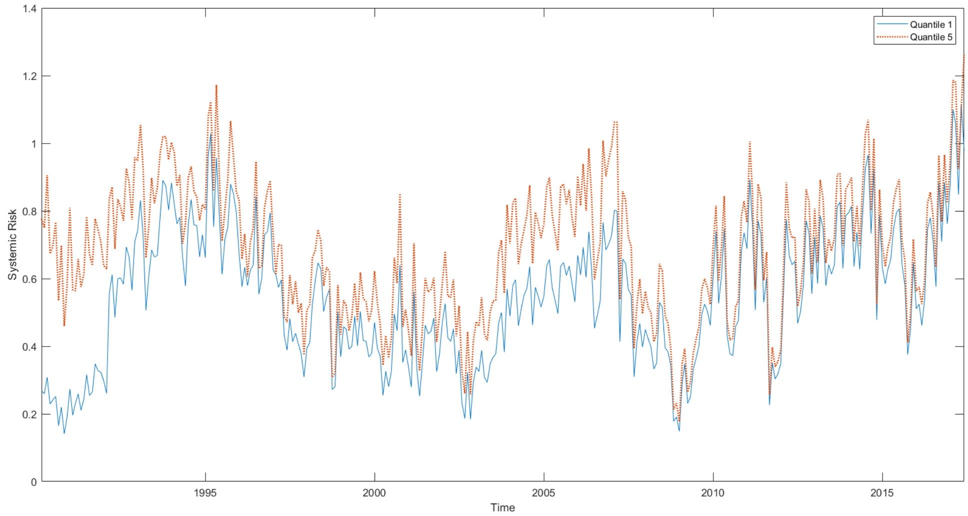

Figure 24: Systemic risk mean of the hedge funds industry weighted
by the AUM, by auto-correlation quantiles

Finally, we observe a positive correlation between the lag 1 auto correlation and the systemic risk coefficient, through both models. That involves that the liquidity exposure would actually increase the link with the system. As panic in the system would tend to involve firesales, hedge funds exposed to more liquidity problems are thus more often affected by these shocks. On the other side, it also plays a role in the size of this shock as the PTR is also impacted. This confirms the important role the auto-correlation plays. We have seen that the auto-correlation lag 1 but also other variables (such as the advance notice or the lockup period) are related to the liquidity exposures. However, the auto-correlation has the strongest theoretical foundation as it has been widely studied and used in contexts comparable to the one of this work. If we separate hedge funds in quantiles depending on the auto-correlation they carry and represent the highest and lowest quantile as in Figure 24, it confirms the positive relationship between the auto-correlation, meaning a proxy of the liquidity exposure, and the systemic risk.

5.4.2 A zoom on the important non-aligned variables

Regarding variables for which the aggregate impact cannot be identified, we can investigate them with a graphical representation, which suffers clearly from a statistical analysis but can give an additional insight. Not surprisingly, if we separate the trend by hedge funds strategy as in Figure 14, the Macro strategy tends to be lower, as it was expected with the previous section. The other strategies however tend to show really analogous behaviours. Hence, in terms of systemic risk, it is really hard to dissociate them as shown in Figure 25.

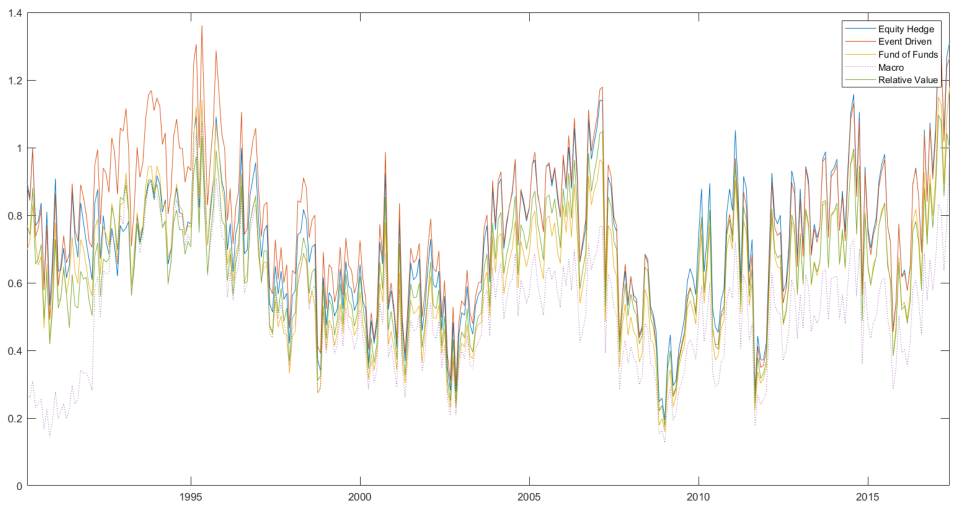

Figure 25: Systemic risk mean of the hedge funds industry weighted by the AUM, by strategy

For the incentive fee, the problem is different, as it can be see on Figure 26. Compared to the end of the period, we have in the beginning of the time horizon a slightly higher systemic risk for the hedge funds with high management fee. For the end of the period, this gap is reversed and seems even bigger. Such a behaviour could be caused by dependence between variables. We might imagine that incentive fee, due to a stronger fear of the market after the crisis, has tended to be more relevant when it comes to the information it provides regarding the quality of the manager. The hedge funds industry might for example realize that there were abusing practices regarding performance compensation of the managers. It thus involved a way too high tail exposure that materialized itself during the crisis, creating this shift.

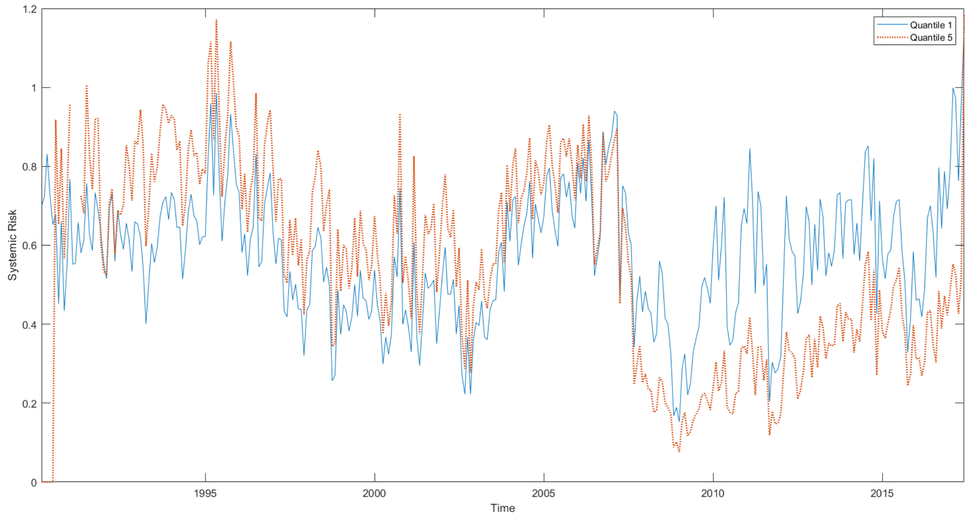

Figure 26: Systemic risk mean of the hedge funds industry weighted
by the AUM, by quantile of incentive fees

The management fee, though the beginning of the period does not show any clear differentiation, the end of the period, after the 2008 crisis, is much clearer as we have significantly higher systemic risk for high management fees.

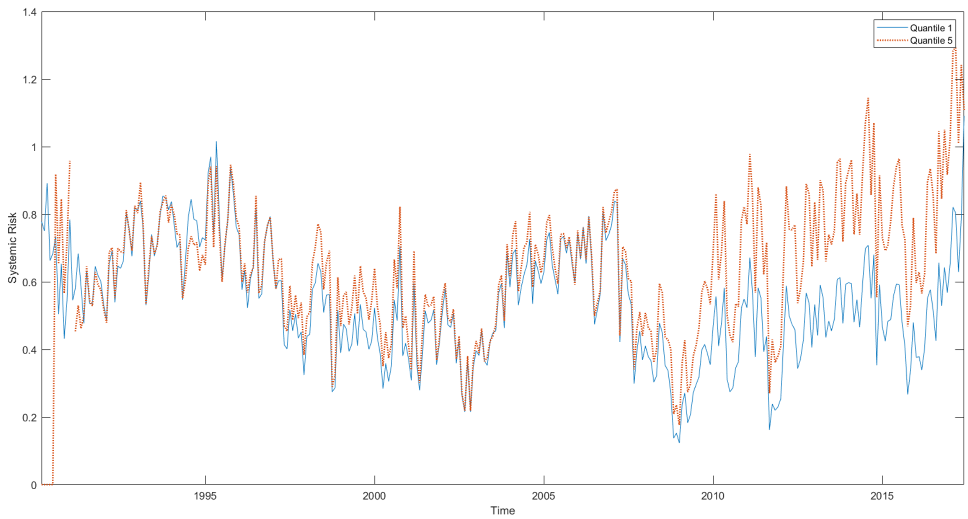

Figure 27: Systemic risk mean of the hedge funds industry weighted
by the AUM, by quantile of management fees

Another potential value we can investigate the same way is the leverage, as illustrated with Figure 28. While the hedge funds using leverage tended to have a higher risk before the crisis, the same, again, does not apply after this financial shock, though it is less clear than the other variables, potentially also due to the narrow range of potential values (as it as dummy variable) and thus information it provides.

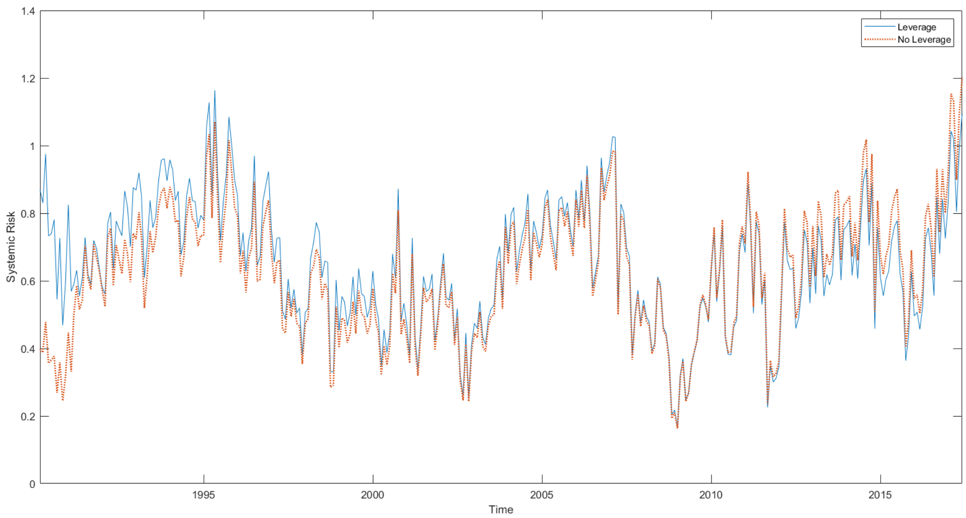

Figure 28: Systemic risk mean of the hedge funds industry weighted by the AUM, depending on the use of leverage

Though we cannot draw any conclusion on the net impact that these three variables have on the systemic risk, it still shows that we have a real change of dynamic after the 2008 crisis. Some variables keep having significant impacts that are consistent with the pre-crisis period, such as the auto-correlation, or the low systemic risk of Macro hedge funds. But others, such as these ones, do not. It would thus require to focus the analysis on this time window and even the most recent data (after 2017).

5.4.3 The final word

As a final set of investigations, given that we wish to have an in-depth assessment of the net impact the PTR and the SL have on the systemic risk, we can find the inspiration in the work of van Oordt and Zhou (2019a), from which the systemic risk model we use comes. Once they estimated the SL and the PTR, the explanation they provide is done using a linear regression against the estimated measures. Thanks to the methodology, the explanation and the measurement are done simultaneously, which allows to overcome the specific challenges of hedge funds in terms of data analysis. Nevertheless, we can also run a linear regression on the final component. This can be seen as a simulation exercise.

Table 9: Linear regression results for the final explanation of the systemic risk coefficient

	Estimate	p-Value
constant	0.701	0.000
Advance Notice	-0.016	0.000
Assets	-0.012	0.000
CPI	-0.001	0.032
EPU	0.020	0.000
High-water mark (yes = 1; no = 0)	-0.001	0.014
Incentive Fee	-0.005	0.000
Ind. Prod.	0.001	0.000
In emerging markets (yes = 1; no = 0)	0.027	0.000
Leverage (yes = 1; no = 0)	-0.003	0.000
Lockup	0.022	0.000
Strategy: Equity Hedge (1)	0.026	0.000
Strategy: Event driven (1)	0.015	0.000
Strategy: Fund of Funds (1)	-0.028	0.000
Strategy: Macro (1)	-0.026	0.000
Management Fee	0.000	0.318
MSCI	0.056	0.000
Offshore Vehicle (yes = 1; no = 0)	-0.026	0.000
Redemptions	0.001	0.000
Negative Lag Returns	0.005	0.000
Unemployment Rate	0.001	0.053
VIX	-0.191	0.000
Auto-correlation 1	0.016	0.000
R^2 adj.	0.384	

Table 9 summarizes these results and shows that we have a net negative impact coming from the incentive fee and the leverage while the PTR and the SL totally offset each other when it comes to the management fee. Though its net impact is not significant, the separate ones on the SL and the PTR are. This means that without this separation, we could have concluded that the management fee does not play a role. As we have seen that the dynamic might have changed after the 2008 crisis, this net impact would have changed too, but would not have been detected to play a significant role precedently. Finally, it confirms again the low systemic risk of Macro hedge funds while it clearly shows that Equity Hedge and Event Driven hedge funds tend to have high systemic risk, meaning that the SL effect tends to dominate the PTR impact. This is actually not the case for the Funds of Funds for which the low PTR seems to mitigate the high tail dependence observed. Once again, we find that the auto-correlation has a significant and positive impact on the systemic risk.

5.5 After 2008?

As we said, the dynamic after 2008 might have changed. Consequently, it can be insightful to run the very same procedure on the time window 2009-2017. This analysis would be beyond the reasonable scope of this work that also suffers from a limitation of the sample size as it uses only one database. Such analysis on such short time interval increases the uncertainty behind the conclusion we could draw. Since it has been shown that the aggregation of database is significantly useful in terms of information but also of data bias management, it might be possible to perform this analysis with higher degree of confidence in such context.

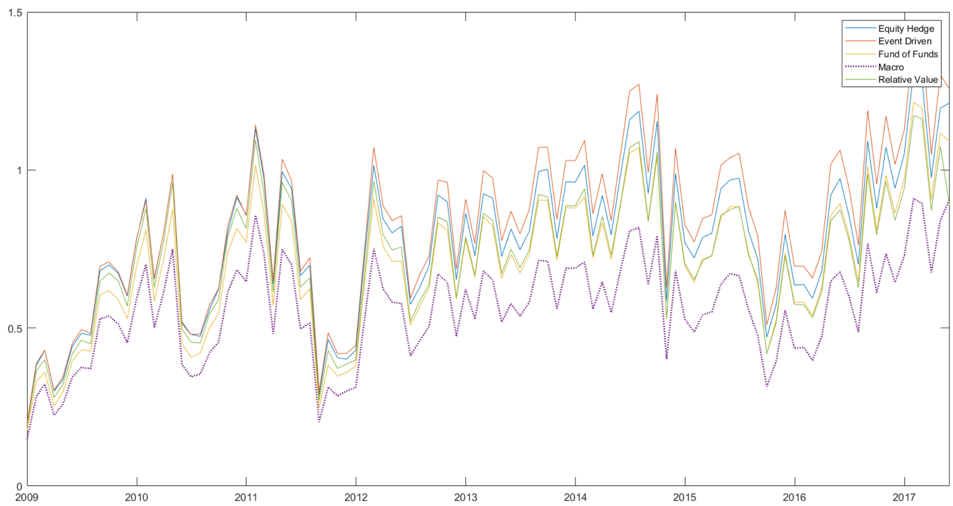

Figure 29: Systemic risk mean of the hedge funds industry weighted by the AUM,grouped by strategy, from the model fitted on the time window 2009-2017

Though this procedure has still been executed, we only discuss the main results for concision. We provide only an illustration of the systemic risk mean over time weighted by the AUM (Figure 29) and the results of the final linear regression of the variables selected by the models against the systemic risk, to gain a certain insight of the net impact on the systemic risk. The main change needed compared to the previous results is that we had to perform another selection based on the lifetime of the individual hedge funds. This selection was needed only for the SL approach due to the presence of hedge funds that were present in the beginning of the time frame but died rapidly after, resulting in really poor fit for the copula and error messages. This selection excluded around 2000 observations for a total sample size of 382298 observations.

The fact that the industry was more developed in terms of existing hedge funds reporting allow us to have a large part of the total sample size inside this time frame, mitigating also the worries pointed before.

First of all, the results confirm that the systemic risk has been increasing during this period. The whole industry is concerned, as Figure 29 illustrates it, but we still find higher and significant coefficients for Relative Value and Equity Hedge hedge funds. The analysis also confirm the actually lower systemic risk of the Macro hedge funds. Moreover, the importance of auto-correlation lag 1 is again selected by the model with a positive impact on the systemic risk. Our conclusion is thus robust to this change. Regarding the coefficients that made us think that a change of dynamic might have happened, the incentive fee is still negatively correlated with the systemic risk of hedge funds. More importantly, we find a negative and significant relationship between the management fee and the final estimation, which was not the case before. The remaining results are consistent with the previous ones, except for the negative lag returns.

Table 10: Systemic risk linear regression results from the model fitted on the time window 2009-2017

	Estimate	p-Value
constant	0.752	0.000
Advance Notice	-0.017	0.000
Assets	-0.011	0.000
CPI	0.025	0.000
EPU	-0.011	0.000
Incentive Fee	0.001	0.002
Ind. Prod.	0.011	0.000
In emerging markets (yes = 1; no = 0)	0.018	0.000
Lockup	0.019	0.000
Strategy: Equity Hedge (1)	0.011	0.000
Strategy: Fund of Funds (1)	-0.026	0.000
Strategy: Macro (1)	-0.035	0.000
Management Fee	-0.001	0.000
MSCI	0.044	0.000
Offshore Vehicle (yes = 1; no = 0)	-0.020	0.000
Redemptions	0.009	0.000
Negative Lag Returns	-0.007	0.000
Unemployment Rate	-0.019	0.000
VIX	-0.162	0.000
Auto-correlation 1	0.022	0.000
R^2 adj.	0.422	

6 Conclusion

As a conclusion, due to the specificities and challenges the hedge funds represent in terms of analysis and modelling, we provide a methodology to apply a systemic risk measurement approach elaborated in the first place for the banking sector. Within this analysis, the methodology allows for a separation of the systemic risk coefficient into two parts: the Pure Tail Risk (PTR) and the Systemic Linkage (SL), a separation that turns out to be useful in our context. It allows notably to detect explanatory variables that would have judged not significant due to the mutual compensation of the two effect, such as the management fee in our case. In addition, the separation enables us to find the methodological solutions to build at the end the total systemic risk coefficient.

On the one side, the LASSO-GP regression provides interesting and significant results, allowing us to give explanations of the PTR but also an analysis of its time-varying behaviour. The results show that classic macro economic variables explain the tail exposures of hedge funds but also that a real differentiation exist between the main strategies with an especially low VaR for the Funds of Funds and Macro hedge funds. Moreover, the analysis points to a negative correlation between the VaR of hedge funds and their size but also with the use of high-water marks. The time-varing PTR, computed through a mean of the industry weighted by the individual size of the funds, evidenced an ability for hedge funds to mitigate their exposures relative to the system, approximated by the S&P500 returns, in times of stress such as the 2008 crisis. The representation by strategy shows that this anticipation effect remains whatever the hedge fund style.

On the other side, the computation and analysis of the SL through the use of copula does not solve entirely the challenges posed by hedge funds and another step is required for the treatment of zero value cases. Not only these approaches solved this problem, the regression tree and the Probit regression enabled to provide the explanatory power that was missing until then, in a relatively complementary way since the regression tree provides a ranking of the criteria. Nevertheless, as lower tail dependence, computed through the Clayton copula, were made constant over the life of each hedge fund, we partially suffer from the lack of a time-varying feature we had with the PTR. However, the results also showed an even clearer difference between the strategies, which is not aligned with the analysis of the PTR.

The last difficulty that this methodology faces is the analysis and computation of the systemic risk coefficient. Since the PTR and the SL are non-linearly combined, our capability to identify what really impacts the systemic risk at the end in terms of explanatory variables and their relative importance is limited but possible. At this time, the impact coming from the PTR and the SL align themselves on the advance notice, the size of hedge funds, the use of high-water mark, the Macro hedge funds and the use of offshore vehicle with a negative correlation with the systemic risk while Lockup period and the auto-correlation are positively linked with it. Moreover, the overall discussion revolved also around the explanation of liquidity risk through the use of the lag 1 auto-correlation in the regression approach. For both the PTR and the SL, the auto-correlation is significant and with an aligned impact, involving an overall positive correlation between the systemic risk and the liquidity exposures of hedge funds. We can find also a certain link with the advance notice, which could give a proxy to liquidity risk reducer and the increased lockup period to be a liquidity risk indicator too. These aspects could be further investigated. In addition, there was a clear distinction between the strategies for both the PTR and the SL, but these do not align with each other for the systemic risk. Nevertheless, the final linear regression enabled us to show that Equity Hedge hedge funds and Event Driven hedge funds are the most subject to high systemic risk. This fact is even more important as the number of the Equity Hedge hedge funds in the sample of the HFR database is the most important.

Driven mostly by the PTR, we can observe that the anticipation effect discussed by Agarwal et al. (2017) is present in the same way for the systemic risk of hedge fund. Thanks to this time-varying feature, we were also able to observe a clear augmentation of the systemic risk after the 2008 crisis. It has also been noticed that the potential dynamic of this particular period (2008-2017) might have changed compared to the past, involving that a particular focus would be required on this time frame.

The approach suffers from limitation and strong assumptions were sometimes needed to carry the analysis until the end, mostly through the analysis of the SL. Moreover, the aggregation of databases would enable to provide more insightful information, notably by having the opportunity to treat with more rigour the information bias the usual hedge funds databases usually carry. However, the results showed strong potential for deeper and wider studies, such as the need to focus on the post-crisis period where our findings stress a potential shift in the tail dynamic of the industry. Besides, extensions of this work are possible for other players in the shadow banking phenomenon and the systemic risk they convey through it, such as the mutual funds. This work thus provides several and interesting paths for other related research topics.

We believe that our efforts to identify and explain the systemic risk of the hedge funds industry and its components generated significant results, showing the increasing importance of a possible regulation. Our findings highlight the real threat that the industry represents and, more particularly, the non-traditional dynamic of it, intrinsic to hedge funds. A key practical element identified by this work revolves around the negative link between the size of one entity and its threat for the financial system. This involves a shift of the regulation perspective in comparison to the banking sector, where bigger players tend to be more monitored than smaller ones.

7 Appendices

7.1 Hedge funds variables description

- numerical variables
  - Assets: the assets under management
  - Returns: returns of the hedge funds
  - Management fee: The management fee of the managers of the hedge funds, an indicator for its remuneration.
  - Incentive fee: The incentive fee is the performance-based remuneration of the manager.
  - Lockup: Time during which a new investor cannot redeem the assets after his or her investment.
  - Redemption: The redemption interval (in days).
  - Advance Notice: The number of days notice required before a redemption.
  - CPI: Consumer Price Index of the referenced country of the hedge fund.
  - Ind. Prod.: Production indices of the referenced country of the hedge fund.
  - Unemployment Rate: The unemployment rate of the country in which the hedge fund is registered.
  - VIX: The implied volatility index, representing the expected volatility of the market as the volatility needed for the Black and Scholes formula to equal the observed prices of an option.
  - MSCI: Equity world index returns.
  - EPU: Economic Policy Uncertainty: index of uncertainty based on current frequencies of words on the web related to uncertainties in the financial market or in the economy.
  - Negative lag returns: Negative return of the previous period. This data could be useful to catch a GARCH-like effect.
- Dummy variables:

– Leverage (No = 0 ; Yes = 1): does the hedge fund use leverage

– High-water mark (No = 0 ; Yes = 1): does the hedge fund use high-water mark, which is a protection for investors against the possibility to remunerate the manager while the performance was poor.

– In emerging market (No = 0 ; Yes = 1): does the hedge fund invest in emerging markets

– Offshore vehicle (No = 0 ; Yes = 1): does the hedge fund use offshore vehicle

7.2 Liquidity model of Getmansky et al. (2003)

In their model, the true return is given by

$$R_t = \mu + \beta\Lambda_t + \epsilon_t \tag{20}$$

with

$$E[\Lambda_t] = E[\epsilon_t] = 0$$

$$Var[R_t] = \sigma^2$$

and

$$\epsilon_t, \Lambda_t \sim IID$$

The idea is that we cannot know the true real return R_t but we can have the observed or reported return R_t^o. They thus assume these observed returns to be

$$R_t^o = \theta_0 R_t + \theta_1 R_{t-1} + ... + \theta_k R_{t-k} \tag{21}$$

With

$$\theta_j \in [0, 1], j = 0, ..., k$$

$$1 = \theta_0 + \theta_1 + ... + \theta_k.$$

They estimate the model (and the θ_j) with a Maximum Likelihood Estimation. With these estimates of $\theta_j, j = 0, ..., k$, we can estimate many quantities related to the illiquidity of hedge funds.

Among their theoretical results, they showed that:

$$Var[R_t^o] = c_\sigma^2 \sigma^2 \leq \sigma^2 \tag{22}$$

where

$$c_\sigma^2 = \theta_0^2 + \theta_1^2 + ... + \theta_k^2$$

But also that the auto-correlation of lag m is

$$Corr[R_t^o, R_{t-m}^o] = \frac{\sum_{j=0}^{k-m} \theta_j \theta_{j+m}}{\sum_{j=0}^{k} \theta_j^2}.$$

We can see two concepts with this theoretical explicitation. First, the auto-correlation depends on the θ_j, which are the parameters of the smooth equation 21. It illustrates the relation between the smoothing and the auto-correlation. Secondly, the variance of the observed returns are lower that the ones of the true returns, involving also that the risk and performance measurement of hedge funds is impacted.

7.3 LASSO-GP Regression by strategy

These tables give the different coefficient estimates of the Lasso-GP regression on datasets grouped by strategies where n gives the total sample of the hedge funds of the mentioned strategy.

Table 11: Lasso GP fit on strategies: Equity Hedge (n=288190) and Event Driven (n=67891)

	Equity Hedge				Event Driven			
	Coefficients σ		Coefficients ξ		Coefficients σ		Coefficients ξ	
	Estimates	p-Value	Estimates	p-Value	Estimates	p-Value	Estimates	p-Value
constant	1.404	0.000	0.102	0.000	1.229	0.000	0.196	0.000
Assets	-0.033	0.000						
Management Fee	0.018	0.064						
Incentive Fee	0.025	0.014						
Lockup	0.071	0.000						
Redemptions	0.017	0.105						
Advance Notice	-0.016	0.105						
CPI	0.077	0.000						
Ind. Prod.	-0.012	0.181						
Unemployment Rate	-0.120	0.000						
VIX	0.206	0.000	-0.028	0.060	0.142	0.000		
MSCI	0.098	0.000	-0.071	0.000	0.150	0.000		
EPU	0.086	0.000	-0.047	0.000	0.122	0.000		
Negative Lag Returns	0.226	0.000	-0.021	0.009				
Auto-correlation 1	0.032	0.010						
High-water mark (yes = 1; no = 0)	-0.025	0.012						
In emerging markets (yes = 1; no = 0)	0.054	0.000						
Leverage (yes = 1; no = 0)	0.014	0.134						
Offshore Vehicle (yes = 1; no = 0)	-0.048	0.000						

Table 12: Lasso GP fit on strategies: Fund of funds (n=199891) and Macro (n=124923)

	Fund of Funds				Macro			
	Coefficients σ		Coefficients ξ		Coefficients σ		Coefficients ξ	
	Estimates	p-Value	Estimates	p-Value	Estimates	p-Value	p-Value	p-Value
constant	0.987	0.000	-0.041	0.000	1.286	0.000	0.097	0.000
Assets	-0.040	0.000						
Management Fee	0.019	0.054						
Incentive Fee	0.056	0.000			0.111	0.000		
Lockup	0.027	0.010			0.046	0.001		
Redemptions	-0.084	0.000						
Advance Notice	-0.030	0.007						
CPI	-0.022	0.050						
Ind. Prod.	0.019	0.057						
Unemployment Rate	-0.053	0.000						
VIX	0.217	0.000			0.126	0.000	0.031	0.093
MSCI	0.337	0.000	-0.142	0.000			0.038	0.002
EPU	0.053	0.000						
Negative Lag Returns	0.150	0.000						
Auto-correlation 1	-0.062	0.000			0.052	0.000		
High-water mark (yes = 1; no = 0)								
In emerging markets (yes = 1; no = 0)	0.062	0.000						
Leverage (yes = 1; no = 0)	0.072	0.000					0.072	0.000
Offshore Vehicle (yes = 1; no = 0)	0.004	0.684			-0.074	0.000	-0.031	0.031

Table 13: Lasso GP fit on strategies: Relative value (n=114408)

	Relative Value			
	Coefficients σ		Coefficients ξ	
	Estimates	p-Value	Estimates	p-Value
constant	1.196	0.000	0.165	0.000
Assets	-0.076	0.000		
Management Fee	0.117	0.000		
Incentive Fee	0.134	0.000		
Lockup	-0.024	0.156		
Redemptions	-0.105	0.000		
Advance Notice	-0.019	0.301		
CPI	0.058	0.001		
Ind. Prod.	-0.069	0.000		
Unemployment Rate	-0.081	0.000		
VIX	0.244	0.000		
MSCI	0.158	0.000	-0.083	0.000
EPU	0.123	0.000		
Negative Lag Returns	0.287	0.000	-0.072	0.000
Auto-correlation 1	0.020	0.257		
High-water mark (yes = 1; no = 0)	-0.052	0.007		
In emerging markets (yes = 1; no = 0)	0.014	0.402		
Leverage (yes = 1; no = 0)	0.015	0.363		
Offshore Vehicle (yes = 1; no = 0)	-0.063	0.000		

7.4 GARCH fit for EVT-GARCH of the S&P500

Table 14: Summary GARCH(1,1)

	Value	p-Value
Constant	0.000	0.000
GARCH1	0.896	0.000%
ARCH1	0.093	0.000%
LLF	2.606e+04	

Table 15: Summary GARCH(2,1)

	Value	p-Value
Constant	0.000	0.000
GARCH(1)	0.876	0.000
GARCH(2)	0.000	1.000
ARCH(1)	0.105	0.000
LLF	2.606e+04	

Table 16: Summary GARCH(1,2)

	Value	p-Value
Constant	0.000	0.000
GARCH(1)	0.884	0.000
ARCH(1)	0.068	0.000
ARCH(2)	0.034	0.001
LLF	2.607e+04	

Table 17: Summary GARCH(2,2)

	Value	p-Value
Constant	0.000	0.000
GARCH(1)	0.213	0.072
GARCH(2)	0.601	0.000
ARCH(1)	0.0653	0.000
ARCH(2)	0.099	0.000
LLF	2.607e+04	

7.5 Split table for the regression tree

Table 18: Split table for the regression tree, Criteria gives the variable that is selected, the SplitPoint gives the value of the criteria to compare with the observation and the Children if lower and bigger respectively gives the Criteria Number to which the observation must be assigned. The process goes on until the row to which the observation is assigned is empty, involving that the group number is a number different from zero that actually gives the group number

Criteria Number	Criteria	Split Point	Children if lower	Children if bigger	Group
1	Strategy: Macro (1)	0.500	2	3	0
2	Incentive Fee	15.688	4	5	0
3	Incentive Fee	13.500	6	7	0
4	Offshore Vehicle (yes = 1; no = 0)	0.500	8	9	0
5	Strategy: Relative Value (1)	0.500	10	11	0
6	Leverage (yes = 1; no = 0)	0.500	12	13	0
7	In emerging markets (yes = 1; no = 0)	0.500	14	15	0
8	Leverage (yes = 1; no = 0)	0.500	16	17	0
9	Advance Notice	5.500	18	19	0
10	Redemptions	37.500	20	21	0
11	Management Fee	1.940	22	23	0
12	Management Fee	0.925	24	25	0
13	Redemptions	18.500	26	27	0
14	Management Fee	1.775	28	29	0
15	Management Fee	1.825	30	31	0
16	Redemptions	4.000	32	33	0
17	Incentive Fee	8.750	34	35	0
18	Incentive Fee	0.820	36	37	0
19	Redemptions	75.000	38	39	0
20	Leverage (yes = 1; no = 0)	0.500	40	41	0
21	Auto-correlation 1	0.091	42	43	0
22	Advance Notice	7.500	44	45	0
23			0	0	1
24			0	0	2
25			0	0	3
26			0	0	4
27			0	0	5
28	Advance Notice	32.500	46	47	0
29	Lifetime	100.500	48	49	0
30			0	0	6
31			0	0	7
32			0	0	8
33			0	0	9
34			0	0	10
35			0	0	11
36			0	0	12
37			0	0	13
38			0	0	14
39			0	0	15
40			0	0	16
41			0	0	17
42			0	0	18
43			0	0	19
44			0	0	20
45			0	0	21
46			0	0	22
47			0	0	23
48			0	0	24
49			0	0	25

References

Acharya, V., Engle, R., & Richardson, M. (2012, May). Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks. *American Economic Review*, *102*(3), 59–64. doi: 10.1257/aer.102.3.59

Adrian, T., & Brunnermeier, M. K. (2016, July). CoVaR. *American Economic Review*, *106*(7), 1705–1741. doi: 10.1257/aer.20120555

Agarwal, V., Bakshib, G., & Huij, J. (2009, August). Do Higher-Moment Equity Risks Explain Hedge Fund Returns? *SSRN Electronic Journal*. doi: 10.2139/ssrn.1108635

Agarwal, V., Ruenzi, S., & Weigert, F. (2017). Tail risk in hedge funds: A unique view from portfolio holdings. *Journal of Financial Economics*, *125*(3), 610–636.

Aragon, G. O., & Nanda, V. (2012). Tournament Behavior in Hedge Funds: High-water Marks, Fund Liquidation, and Managerial Stake. *Review of Financial Studies*, *25*(3), 937–974.

Bali, T. G., Brown, S. J., & Caglayan, M. O. (2014, October). Macroeconomic risk and hedge fund returns. *Journal of Financial Economics*, *114*(1), 1–19. doi: 10.1016/j.jfineco.2014.06.008

Ben-David, I., Franzoni, F. A., & Moussawi, R. (2011, September). *Hedge Fund Stock Trading in the Financial Crisis of 2007-2009* (SSRN Scholarly Paper No. ID 1550240). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.1550240

Bengtsson, E. (2016). Investment funds, shadow banking and systemic risk. *Journal of Financial Regulation and Compliance*, *24*(1), 60–73.

Billio, M., Getmansky, M., & Pelizzon, L. (2012). Dynamic risk exposures in hedge funds. *Computational Statistics & Data Analysis*, *56*(11), 3517–3532.

Billio, M., Getmansky Sherman, M., & Pelizzon, L. (2010, July). *Crises and Hedge Fund Risk* (SSRN Scholarly Paper No. ID 1130742). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.1130742

Boyson, N. M., Stahel, C. W., & Stulz, R. M. (2010). Hedge Fund Contagion and Liquidity Shocks. *The Journal of Finance*, *65*(5), 1789–1816. doi: 10.1111/j.1540-6261.2010.01594.x

Bussière, M., Hoerova, M., & Klaus, B. (2014, February). *Commonality in Hedge Fund Returns: Driving Factors and Implications* (SSRN Scholarly Paper No. ID 2034894). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.2034894

Chan, N., Getmansky, M., Haas, S. M., & Lo, A. W. (2005, March). *Systemic Risk and Hedge Funds* (Tech. Rep. No. w11200). National Bureau of Economic Research. doi: 10.3386/w11200

Chavez-Demoulin, V., Embrechts, P., & Hofert, M. (2016). An Extreme Value Approach for Modeling Operational Risk Losses Depending on Covariates. *Journal of Risk and Insurance*, *83*(3), 735–776. doi: 10.1111/jori.12059

Edelman, D., Fung, W., Hsieh, D., & Naik, N. (2012, March). Funds of hedge funds: Performance, risk and capital formation 2005 to 2010. *Financial Markets and Portfolio Management*, *26*, 87–108. doi: 10.1007/s11408-011-0180-z

Embrechts, P., Haan, L. D., Huang, X., & Instituut, E. (1999). *Modelling Multivariate Extremes* (Tech. Rep.).

Fung, W., & Hsieh, D. A. (2002, January). Hedge-Fund Benchmarks: Information Content and Biases. *Financial Analysts Journal*, *58*(1), 22–34. doi: 10.2469/faj.v58.n1.2507

Fung, W., & Hsieh, D. A. (2004, September). Hedge Fund Benchmarks: A Risk-Based Approach. *Financial Analysts Journal*, *60*(5), 65–80. doi: 10.2469/faj.v60.n5.2657

Fung, W., & Hsieh, D. A. (2009, May). Measurement Biases in Hedge Fund Performance Data: An Update. *Financial Analysts Journal*, *65*(3), 36–38. doi: 10.2469/faj.v65.n3.6

Getmansky, M., Lo, A., & Makarov, I. (2004, May). An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns. *SSRN Electronic Journal*. doi: 10.2139/ssrn.384700

Gregoriou, G. N., Racicot, F.-É., & Théoret, R. (2021, January). The response of hedge fund tail risk to macroeconomic shocks: A nonlinear VAR approach. *Economic Modelling*, *94*, 843–872. doi: 10.1016/j.econmod.2020.02.025

Hambuckers, J., Groll, A., & Kneib, T. (2018). Understanding the economic determinants of the severity of operational losses: A regularized generalized Pareto regression approach. *Journal of Applied Econometrics*, *33*(6), 898–935. doi: 10.1002/jae.2638

Harri, A., & Brorsen, B. (2004). Performance persistence and the source of returns for hedge funds. *Applied Financial Economics*, *14*(2), 131–141.

Hespeler, F., & Loiacono, G. (2017). Monitoring systemic risk in the hedge fund sector. *Quantitative finance*, *17*(12), 1859–1883. doi: 10.1080/14697688.2017.1357969

Implementation of G20/FSB financial reforms in other areas. (2019). , 31.

Joe, H. S. W. (1997). *Multivariate models and dependence concepts /.* London :: Chapman and Hall,.

Joenväärä, J., Kauppila, M., Kosowski, R., & Tolonen, P. (2019, January). *Hedge Fund Performance: Are Stylized Facts Sensitive to Which Database One Uses?* (SSRN Scholarly Paper No. ID 1989410). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.1989410

McNeil, A. J., & Frey, R. (2000, November). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, *7*(3-4), 271–300. doi: 10.1016/S0927-5398(00)00012-8

Mhalla, L., Hambuckers, J., & Lambert, M. (2020, July). *Extremal Connectedness and Systemic Risk of Hedge Funds* (SSRN Scholarly Paper No. ID 3519295). Rochester, NY: Social Science Research Network. doi: 10.2139/ssrn.3519295

Mikosch, T., & Starica, C. (2000). Limit Theory for the Sample Autocorrelations and Extremes of a GARCH (1, 1) Process. *The Annals of Statistics*, *28*(5), 1427–1451.

Racicot, F.-É., & Théoret, R. (2016). Macroeconomic shocks, forward-looking dynamics, and the behavior of hedge funds. *Journal of Banking & Finance*, *62*(C), 41–61.

Teo, M. (2011). The liquidity risk of liquid hedge funds. *Journal of Financial Economics*, *100*(1), 24–44.

van Oordt, M., & Zhou, C. (2019a). Estimating Systematic Risk under Extremely Adverse Market Conditions. *Journal of Financial Econometrics*, *17*(3), 432–461.

van Oordt, M., & Zhou, C. (2019b). Systemic risk and bank business models. *Journal of Applied Econometrics*, *34*(3), 365–384. doi: 10.1002/jae.2666

Executive Summary

This master's thesis studies the systemic risk of the hedge funds industry through a model developed and used by van Oordt and Zhou (2019a, 2019b). This model measures systemic risk by splitting it into two components: the Pure Tail Risk, defined as the relative tail exposure towards the whole financial system approximated by the S&P500, and the Systemic Linkage, which represents the probabilistic link of the extreme negative returns between the hedge funds and that proxy.

Our first intended contribution resides in the explanation and the reconciliation of these two measures, the downside risk-related component and the tail dependence aspect, into one single metric. Our second contribution relates to innovations brought in the estimation method, which relies on Extreme Value Theory to overcome the scarcity of data and the low frequency of reporting inherent to the hedge funds databases. In order to do so, we implement a LASSO-Generalized Pareto Regression for tail exposure explanations and use copula distributions for tail dependence measurement.

Our work provides several new insights. We observe a significant positive correlation between liquidity and systemic risk indicators, which highlights the link between shadow banking, hedge funds and systemic risk. Moreover, a larger fund size and a longer advance notice act as systemic risk reducers, while the lockup period drives up the systemic threat. Driven by the Pure Tail Risk, the results show an increase of the systemic risk during period of stability such as the period of 2003 to 2006 and after the 2008 crisis. Coupled with the high commonality of the industry in such periods, observed by Bussière et al. (2014) between 2003 and 2006, our findings highlight the potential threat that hedge funds bring to the stability of the financial system. We uncover a common dynamic behaviour of hedge fund strategies over time, even though Equity Hedge and Event Driven hedge funds show significantly higher values. This results is all the more important given that we observe that the Equity Hedge strategy represents the largest fraction of the industry. Finally, our model stresses a shift of the hedge funds tail risk dynamic after the 2008 crisis, showing a genuine potential for future research.

As a result, we believe that our efforts to identify and explain the systemic risk of the hedge funds industry and its components have led to insightful results, which point to an increasing need for an adapted regulation.