

**Before and After the Financial Crisis:
Reevaluating Hedge Fund Survival Probability Model**

Joseph Wharton Scholar Program

WH399 Thesis

Tianni Lin

Faculty Advisor: Christopher Geczy

April 27, 2016

I. Introduction

The hedge fund industry has grown substantially in the past two decades. In 2012, it reached a new record of more than two trillion dollars of assets under management. Many scholars have previously studied the survival models of hedge funds; however, the industry evolves quickly, and it is possible that the works published more than a decade ago might not be relevant today. In addition, given few has studied the survival models by using variables such as minimum investment, investment strategy, leverage, fund asset, management fee, incentive fee, high watermark, redemptions, lockup period, diversity firm, SEC registration, performance, and advance notice after the 2008 financial crisis, I would like to dedicate my research to fill the gap. Specifically, this research paper will examine the relations between hedge fund survival probability and the variables mentioned above, taking data after the financial crisis into account.

The rest of this research proposal is organized as follows. Section II conducts a literature review. Section III presents the significance of the research. Section IV lays out the hypotheses, and sections V and VI explain the data and the methodology, respectively. Last but not the least, sections VII and VIII will show and discuss the results, followed by section IX conclusion in the end.

II. Literature Review

The topic of hedge fund lifetimes has been well researched. Using TASS database, Brown *et al.* (2001) find that hedge fund termination is a function of performance relative to the rest of the industry. They also discover that hedge funds have a half-life survival of 2.5 years. In particular, young funds die faster, which is a finding confirmed by Amin and Kat (2002). Howell (2001) observed that the probability of hedge funds failing in their first year was 7.4%, only to increase

to 20.3% in their second year. This is easy to understand in the sense that new funds with a small amount of asset under management are likely to take on additional risk to enhance their returns and attract investors. Rather than using their reputation, they prefer to use performance to solicit investors. Brooks and Kat (2001) detect that approximately 30% of new hedge funds do not make it past 36 months owing to poor performance. Young funds that perform badly drop out of databases at a faster rate than older funds according to a study examining 3,477 live and dead funds during the period 1980–2000 (Jen *et al.*, 2001).

In addition, volatility and manager tenure were used as covariates. Brown *et al.* (2001) also finds that the attrition rate in hedge funds since 1994 was almost 15% per year. Furthermore, they find evidence that the probability of death increases with increasing risk, and that funds with negative returns for two consecutive years have a higher risk of shutting down. Once again, they group all hedge fund styles in aggregate, rather than investigating each classification separately.

Moreover, Amin and Kat (2002) find that size and leverage have an effect on the survival of hedge funds and discovered that more than 40% of hedge funds do not make it to their fifth year, and the ones that did represented only 60% of the sample, confirming the results of Liang (2000). Boyson (2002) reaches the same conclusion, and observes that hedge fund size has a strong effect on survival that the smaller funds die faster than the larger ones. In the same vein, using the Zurich Capital Markets database, Gregoriou (2002) examines several predictors, and demonstrates that some variables can be useful to predict hedge fund mortality. It is found that size, along with redemption period, performance fee, leverage, monthly returns and minimum purchase have a significant correlation with the mortality of hedge funds, and this correlation holds the true for both live and defunct funds. More recent work by Gregoriou *et al.* (2009) shows that exchange-listed hedge funds typically are larger and more conservative than non-

listed funds. Compared with non-listed funds, exchange-listed hedge funds perform more poorly while having lower volatility. Listed hedge funds appear to survive about two years longer than non-listed funds.

Some scholars also approach the survival analysis from unique angles. Bares *et al.* (2001) use a genetic algorithm and observe that optimal portfolio weights are influenced by survival times, allowing the authors to compare multiple prediction models. They also discover that hedge fund managers cease their operations without any prior notification as a result of extreme market events from which they do not expect to recover and continue operations. Boyson (2002) examines manager age, tenure, and SAT scores and finds a negative relationship between these variables and the likelihood of failure. Lee and Yao (2015) investigate the failure probabilities of hedge funds, by using both a proportional hazard model and a logistic model. Using a signal detection model and a relative operating characteristic curve as the prediction accuracy metrics, they find that both models have predictive power in the out-of-sample test. The proportional hazard model, in particular, has stronger predictive power on average.

On a side note, numerous studies have also examined what correlates with hedge fund's performance, which is linked to hedge fund's survival (Brooks and Kat, 2001). Gregoriou and Rouah (2002) examine the returns of hedge funds and funds of hedge funds to determine whether the size of a fund affects its performance, and their findings suggest that the size of a hedge fund or a fund of hedge funds has no impact on its performance. Getmansky (2004) find that there is a concave relationship between performance and lagged assets under management. The implication of this study is that an optimal asset size can be obtained by balancing the effects of past returns, fund flows, competition, market impact, and favorable category positioning that are modeled in the paper. Hedge funds in capacity constrained and illiquid categories are subject to

high market impact, have limited investment opportunities, and are likely to exhibit an optimal size behavior. Other factors that affect the lifetime of funds include lock-up period, redemption frequency, and downside risk measures. Simon (2011) investigates the causal effect of fund lock-up period and fund lifetime and shows that the lock-up period variable should be treated endogenously. He claims that the lock-up period has protection power for funds in case of poor performance. Liang and Park (2010), in their analysis of predicting hedge fund failure, find that average hedge fund returns are related positively to incentive fees, fund assets, and the lockup period.

III. Research Significance

My paper researches on the hedge fund's survival probability in relation to the fund's asset, minimum investment requirement, investment strategy, leverage, management fee, incentive fee, high watermark, redemptions, lockup period, diversity, SEC registration status, performance, and advance notice requirement. Although there have been abundant research on hedge fund's survivorship in the past, none of them has reexamined the issue by taking the 2008 financial crisis into consideration.

The audiences of my research paper are scholars that are interested in hedge funds, hedge fund managers, investors, hedge fund consultants (e.g. investment banks), as well as regulators. Beyond filling the gap in academic research, this research will generate three broader applications. The practical application of this paper is 1) to help hedge fund managers and scholars develop a better understanding of whether hedge fund's survival is correlated with the attributes mentioned above, if taking the 2008 financial crisis into consideration, 2) to help hedge fund consultants as well as investors understand what kind of fund might have higher survival

probability and how they should adjust their decision-making strategy accordingly, and 3) to help regulators better understand the hedge fund industry.

First, the research findings may help hedge fund managers as well as scholars who are interested in capital markets understand hedge fund's survivorship. In reality, due to the management fee scheme, hedge fund managers almost always have incentives to raise more capital in order to charge more management fee as an important source of the fund's income. However, it is possible that larger funds may lead to more difficulties on operating and investing sides of the business. On the other hand, it is possible that smaller asset under management potentially attracts fewer institutional investors as they may question why the fund manager is unable to introduce more capital into the fund.

Second, it will help institutional investors make better judgment when evaluating which funds to invest in. In most cases, institutional investors will tend to pay more attention to investment strategy, portfolio manager's track record, investment philosophy, research method, and risk management scheme, and few would take a technical standpoint to analyze fund's survival probability. However, the survival probability will surely help investors make better judgments when they choose a hedge fund. The study will help them navigate the survival probability of hedge funds, particularly by taking financial distress into consideration.

By the same token, hedge fund consultants (e.g. investment bank's prime brokerage) will also benefit from the research. Many investment banks provide consulting services to hedge fund clients through the prime brokerage arm for free; hedge funds, in turn, bring in revenue for the investment banks by trading and/or financing through the relationship banks. Therefore, it is important for investment banks to pick the funds have the most potential to generate revenue for the banks in the future, and in this regard, survivorship becomes a very important factor in

making the decision. My research will help the hedge fund consultants to interpret whether hedge fund's survival is correlated to the factors being investigated in the research.

Last but not least, from a regulatory perspective, the research will also help regulators understand the incentive of a hedge fund's certain action. After the financial crisis, regulators have exerted extra amount of effort to promote better governance in the financial market. Therefore, it is helpful for the regulators to have a better understanding of the hedge fund industry. The result will be useful for them to monitor the fund's survivorship in real-time.

IV. Hypothesis

Concretely, I will test two sets of hypotheses in this research:

- whether some variables, individually and/or collectively, have statistically significant effects on the fund's survival probability;
- whether the survival probability model has changed after the 2008 financial crisis.

V. Data

This paper employs the Hedge Fund Research (HFR) database that includes funds that are US-based and USD-denominated. The biases present in hedge fund databases are well documented in literature, and I take measures to mitigate them (see Liang 1999; Fung and Hsieh 2000). As I include both live and defunct funds in the analysis, survivorship bias is reduced. Out of the 65 available variables in the database, I choose 13 variables that may be important to a hedge fund's survivorship.

Here is a list of all the attributes with definition and remark attached, if applicable, that are included in my datasets:

| | Definition | Remark, if applicable |
|---------------------|---|---|
| Minimum Investment | Minimum investment for the fund | |
| Main Strategy | Primary investment strategy | |
| Leverage | Specifies if fund intends to use leverage | |
| Fund Asset | Total assets under Money Management firm | |
| Management Fee | Annual management fee percentage | |
| Incentive Fee | Annual incentive fee percentage | |
| High Watermark | Specifies if fees are taken only after a high watermark (marked as yes or no) | Defined either yes or no |
| Redemptions | Redemption intervals from the fund (i.e. monthly, quarterly, annually) | Defined by number of days |
| Lockup | Lockup interval (length of time that new investor cannot redeem assets) | Defined by number of days |
| Is Diversity Firm | Whether the fund belong to a diversity firm (i.e. women or minority-owned) | Defined by either yes or no |
| Firm SEC Registered | Whether the fund is SEC registered | Defined by either yes or no. Used as a proxy for fund's size as funds that are larger than 100 million USD are required to be registered with the SEC |
| Performance | Historical rate of return | Note: I tuned this variable into geometric average return of historical performance |
| Advance Notice | Indicates advance notice, in days, required for Redemptions | Defined by number of days |

Time Period

This research uses data from two different time periods: September 2002 to October 2007, and March 2009 to April 2014. For the first time period, the reason to choose September 2002 as the starting time is to minimize the influence from dot-com bubble and stock market downturn, during which September 2002 is the lowest of point. I choose October 2007 as the end of the first period because S&P 500 peaked at that time. Therefore, 62 months of data are included in the analysis of the pre-crisis period. For the second time period, I choose March 2009 as the beginning period because the stock market started to pick up at that time. To make two time periods having the same duration, I set April 2014 as the end of the second period, which makes both periods 62 months.

Cleaning and standardizing data

First, I exclude the funds that do not have complete information from the data. Second, for those funds that have complete information but do not have concrete value for each variable -- showing as a range of value (e.g. management fee for certain funds is from 1 to 1.5), I take the average of the range to standardize the variables. Third, since lockup and redemption periods are recorded by using description rather than numbers, I convert all the associated information from verbal description to number of days. After these three steps, in the first dataset, number of fund is downsized from 5609 to 5390, while in the second dataset, from 4675 to 4621.

In addition, I regroup the data for “leverage.” Since there were 11 categories under the “leverage” variable, I regrouped them into 5 levels: 1) no leverage, 2) 1-2 times of leverage, 3) 2-5 times of leverage, 4) 5-10 of times, and 5) more than 10 times of leverage, which are respectively labeled as leverage0, leverage1, leverage2, leverage3, and leverage4. However, since there are only 7 and 26 data points for leverage3 in the first period and the second period

respectively and only 3 and 9 data points for leverage4 in the first period and the second period respectively, leverage3 and leverage4 are therefore deleted because number of data points available is too small.

Moreover, calculation for transferring monthly data to geometric average return is also needed in the data standardizing process. In the original database, historical rate of return is recorded monthly. Therefore, based on the information provided, I took the geometric average of these monthly return of each hedge fund as the indicator for its historical performance.

VI. Methodology

Since our goal is to build a model to understand what variables are important to a hedge fund's survival probability, and what may have changed before and after the 2008 crisis with regard these variables, our final model is a standard logistic regression:

$$p(y = 1|\mathbf{x}; \theta) = \sigma(\theta^\top \mathbf{x}) = \frac{1}{1 + \exp(-\theta^\top \mathbf{x})}.$$

Here each $\mathbf{x}^{(i)} \in R^N$ is an N-dimensional feature vector, where N is the number of variables in the final model; $y^{(i)} \in \{0, 1\}$ is a class label, with 0 showing the fund is dead, and 1 otherwise (i.e. alive); $\theta \in R^N$ are parameters of the logistic regression model; and $\sigma(\cdot)$ is the sigmoid function. With this model fitted on both pre- and post-2008 crisis data, we could draw some inferences from there.

Given the cleaned data, the next step is do feature selection. Two approaches stand out: the modern approach called LASSO, where it uses L_1 regularization that controls the amount of shrinkage in the model to prevent overfitting. It automatically sets parameters of unimportant variables to 0:

$$\min_{\theta} \sum_{i=1}^M -\log p(y^{(i)} | \mathbf{x}^{(i)}; \theta),$$

subject to $\|\theta\|_1 \leq C$.

In order to achieve best possible performance (higher accuracy, or minimum misclassification error), it is necessary to tune “lambda”, or $1/C$ in the above formulation, through methods such as 10-fold cross validation.

Another, traditional approach, is to look at metrics of model output: Adjusted R-Square, Cp, and BIC:

- Adjusted $R^2 = 1 - \frac{\text{RSS}/(n - d - 1)}{\text{TSS}/(n - 1)}$
- $C_p = \frac{1}{n} (\text{RSS} + 2d\hat{\sigma}^2)$
- $\text{BIC} = \frac{1}{n} (\text{RSS} + \log(n)d\hat{\sigma}^2)$

Where d is the number of predictors, and $\hat{\sigma}^2$ is an estimate of the variance of the error ϵ associated with each response. Essentially, the Cp statistic adds a penalty of $2d\hat{\sigma}^2$ to the training RSS in order to adjust for the fact that the training error tends to underestimate the test error. Like Cp, the BIC will tend to take on a small value for a model with a low test error, and so generally we select the model that has the lowest BIC value.

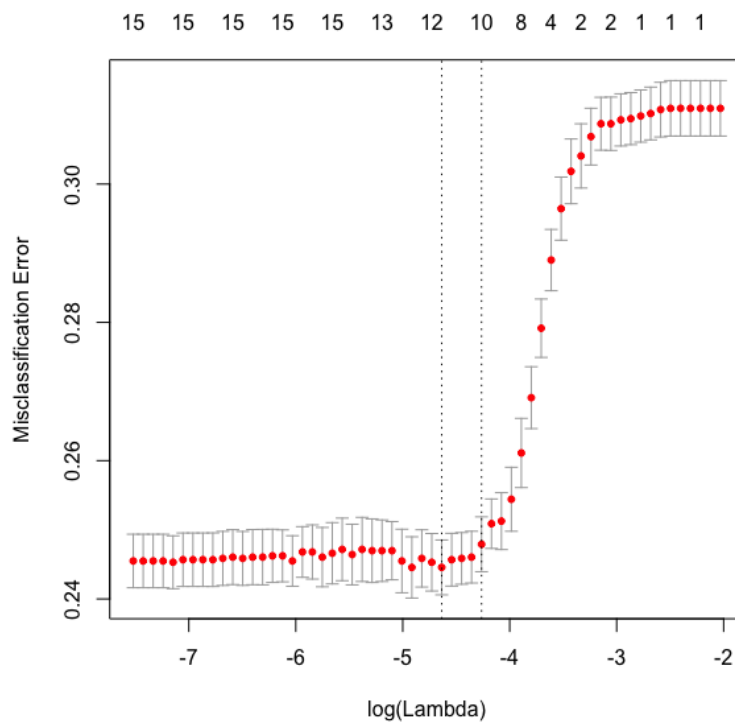
To ensure the variables used in the logistic regression are significant enough, I will only take those variables that are considered significant within the specific time period by both modern and traditional approaches. In addition, certain variables might need to be regrouped (recoded) into other variables within the same category of any of the levels drop from selected variables (more details will be given in the next section). Next, I will put the remaining variables into logistic regression in order to do inference (e.g. p-value).

VII. Results

Pre-crisis: 2002-2007

Modern Approach - LASSO Model

Followed by the methodology mentioned in the previous section, I start with LASSO model to identify significant variables, and here is the output from cross validation tuning the parameter lambda:



LASSO variable output for pre-crisis period:

```

> coef(lasso.cv, s = "lambda.min")
17 x 1 sparse Matrix of class "dgCMatrix"
                                1
(Intercept)                    3.505412e-01
(Intercept)                      .
minimum_investment              -2.865747e-08
main_strategyEvent-Driven       .
main_strategyFund of Funds      -4.266328e-02
main_strategyMacro              -2.951153e-01
main_strategyRelative Value     .
leverage1                       3.817292e-02
leverage2                      -1.291393e+00
management_fee                 -5.992537e-02
incentive_fee                  1.166862e-02
high_watermarkYes              .
advance_notice                 -2.270900e-03
lockup                          .
is_diversity_firmYes           -6.471481e-01
firm_sec_registeredYes         -1.109990e+00
geometric_return               -4.914024e+01

```

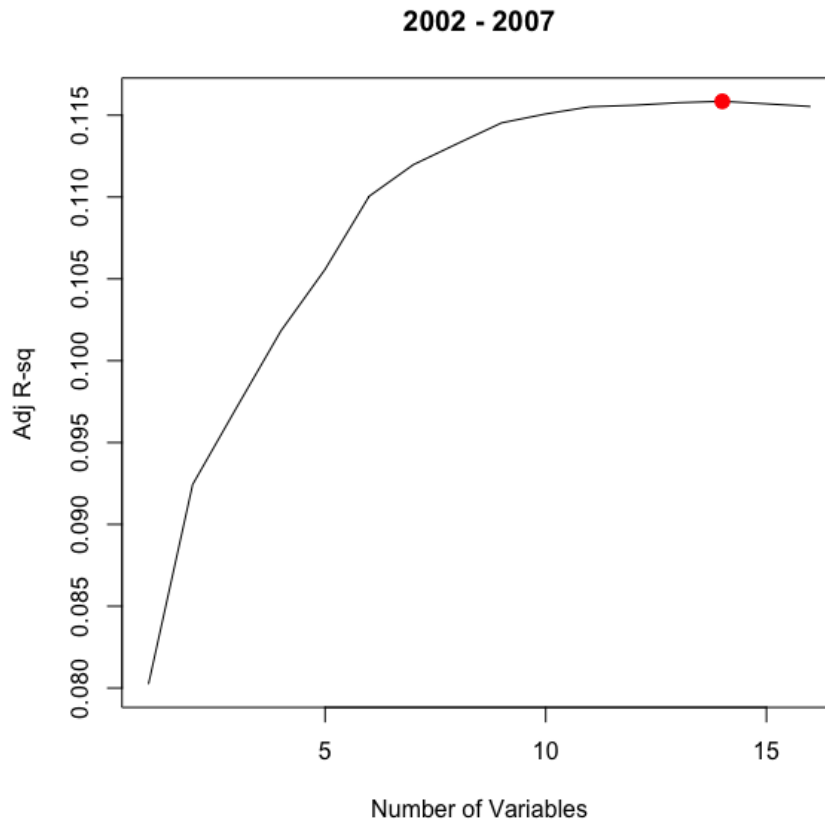
In this table, if a variable is not followed by a number, then it is not selected in the regularized logistic regression by the LASSO model. From this output, we notice that lockup period, high water mark, event-driven strategy, and relative value strategy are considered as not statistically significant by the modern approach.

Traditional Approach

Criterion 1: Adjusted R-Squared

Adjusted R-squared measures the proportion of the variation in the dependent variable accounted for by the explanatory variables, with a model size penalty. By plotting the graph of adjusted R-squared and number of variables included in the logistic regression, we find that the highest adjusted R-squared value corresponds to a 14-variable model. This model indicates that high

watermark should be excluded from the best model.



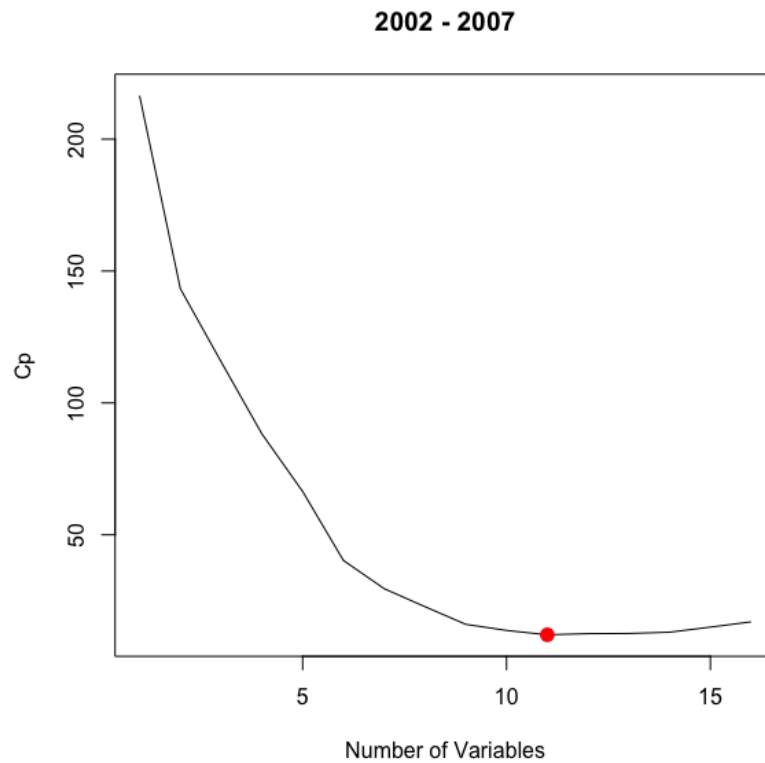
```
> coef(regfit.full,14)
```

| | | | | | |
|----------------------------|---------------|------------------------|---------------|-----------------------------|---------------|
| (Intercept) | 1.579099e+00 | minimum_investment | -5.280418e-09 | main_strategyEvent-Driven | -3.399688e-02 |
| main_strategyFund of Funds | -2.890996e-02 | main_strategyMacro | -1.126155e-01 | main_strategyRelative Value | -3.276076e-02 |
| leverage1 | 2.394563e-02 | leverage2 | -2.334609e-01 | management_fee | -3.642181e-02 |
| incentive_fee | 2.961418e-03 | advance_notice | -9.035102e-04 | lockup | -7.190717e-05 |
| is_diversity_firmYes | -1.586753e-01 | firm_sec_registeredYes | -2.635380e-01 | geometric_return | -1.491121e+00 |

Criterion 2: Cp

In statistics, Mallows's Cp is applied in the context of variable selection, and a small value of Cp means that the model is relatively precise.

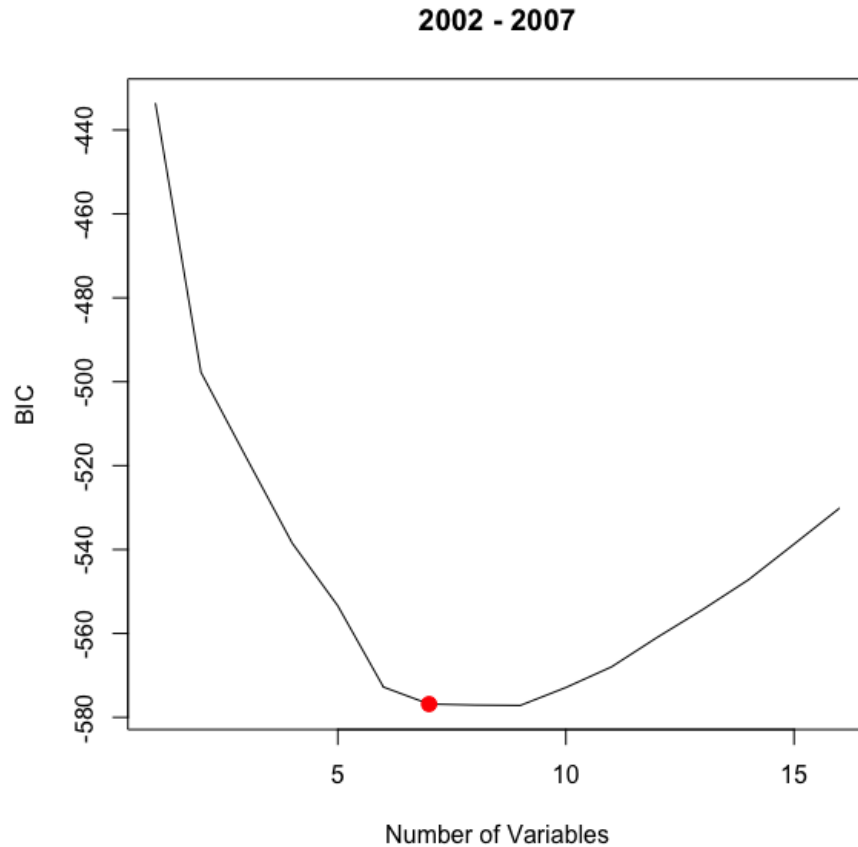
The graph below indicates that when the number of variables equals to 11, Cp has the lowest value. Followed by this conclusion, we pick 11 significant variables, in which high watermark, event driven, fund of funds, and relative value strategies are not included.



```
> coef(regfit.full,11)
(Intercept)      minimum_investment      main_strategyMacro      leverage1
1.562077e+00      -5.404779e-09      -1.012072e-01      2.430876e-02
leverage2        management_fee      incentive_fee      advance_notice
-2.370359e-01      -3.739912e-02      3.529683e-03      -1.054242e-03
lockup           is_diversity_firmYes      firm_sec_registeredYes      geometric_return
-6.541522e-05      -1.607482e-01      -2.656720e-01      -1.478595e+00
```

Criterion 3: BIC (Bayesian information criterion):

According to the output of BIC model selection, only 7 variables should be included in the logistic regression. The selected variables are advance notice, macro strategy, diversity, leverage2, SEC registration, incentive fee, as well as geometric return.



```
> coef(regfit.full,7)
      (Intercept)      main_strategyMacro      leverage2      incentive_fee
      1.520679970      -0.111102048      -0.251430218      0.003377803
  advance_notice  is_diversity_firmYes  firm_sec_registeredYes  geometric_return
      -0.001212861      -0.167943378      -0.269415828      -1.527530682
```

Based on the outcome from both modern and traditional variable selection approaches, there are two leverage levels (leverage0 and leverage1) that are significant enough, and I recode leverage2

into “high” leverage, and combine leverage0 and leverage1 into “low” leverage. Similarly, only “macro” level in strategy is statistically significant from other levels, and thus I recode all other levels in strategy into “else”, and only “macro” as an individual variable.

Applying the modern and traditional variable selection approaches and regrouping variables, I put the remaining variables into logistic regression to do inference (e.g. p-value). From the output, indeed all variables are significant at 0.01 level.

```
> summary(fit)

Call:
glm(formula = dat$fund_status ~ ., family = binomial, data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-4.0385  -0.7734  -0.5799   0.9428   4.7029

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.127355   0.146475  -0.869  0.38459
main_strategyelse  0.658209   0.103828   6.339 2.31e-10 ***
leveragehigh     -2.849364   0.623801  -4.568 4.93e-06 ***
incentive_fee     0.023088   0.005387   4.286 1.82e-05 ***
advance_notice    -0.004362   0.001327  -3.287 0.00101 **
is_diversity_firmYes -1.087426   0.219632  -4.951 7.38e-07 ***
firm_sec_registeredYes -1.259839   0.067394 -18.694 < 2e-16 ***
geometric_return  -80.762970   4.641564 -17.400 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

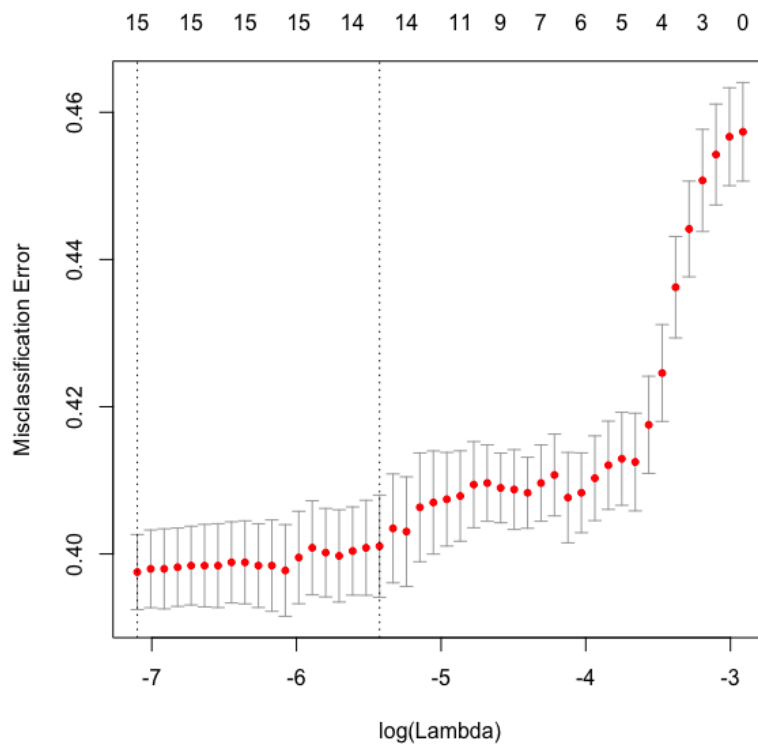
    Null deviance: 6666.0  on 5376  degrees of freedom
Residual deviance: 5646.2  on 5369  degrees of freedom
AIC: 5662.2

Number of Fisher Scoring iterations: 6
```

Post-crisis: 2009-2014

Modern Approach - LASSO Model:

I follow the same steps used to analyze the previous period, and graphs below indicate the output from LASSO model for the period of 2009-2014.

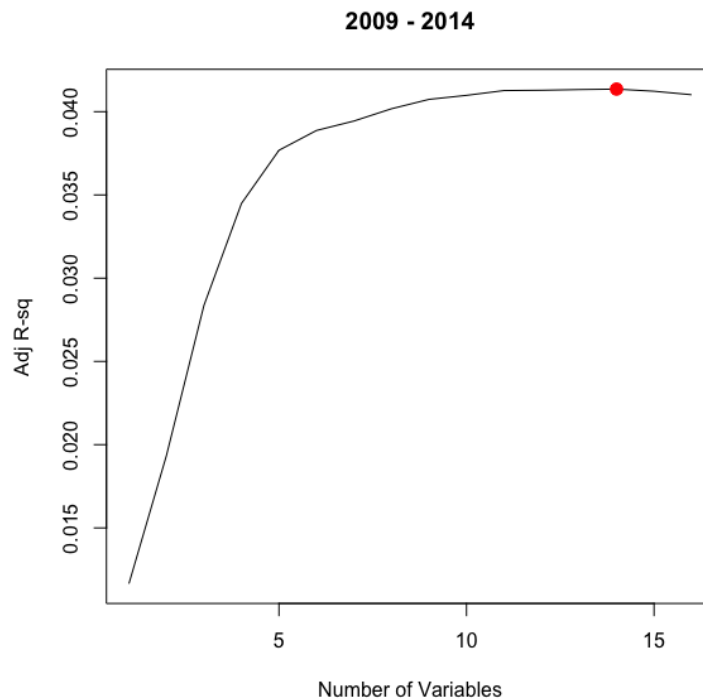


```
> coef(lasso.cv, s = "lambda.min")
17 x 1 sparse Matrix of class "dgCMatrix"
1
(Intercept) -8.849026e-01
(Intercept) .
minimum_investment -2.499300e-08
main_strategyEvent-Driven -1.683108e-01
main_strategyFund of Funds 7.424358e-01
main_strategyMacro -2.396073e-01
main_strategyRelative Value -7.952825e-02
leverage1 -8.804618e-02
leverage2 -3.340117e-01
management_fee 2.844984e-01
incentive_fee 4.425121e-02
high_watermarkYes 1.630316e-02
advance_notice -1.713673e-03
lockup 6.286217e-05
is_diversity_firmYes -1.089547e-01
firm_sec_registeredYes -3.248947e-01
geometric_return -1.586193e+01
```

Traditional methods:

Criterion 1: Adjusted R-Squared

According to the adjusted R-squared criterion, high watermark is eliminated from the model selection.



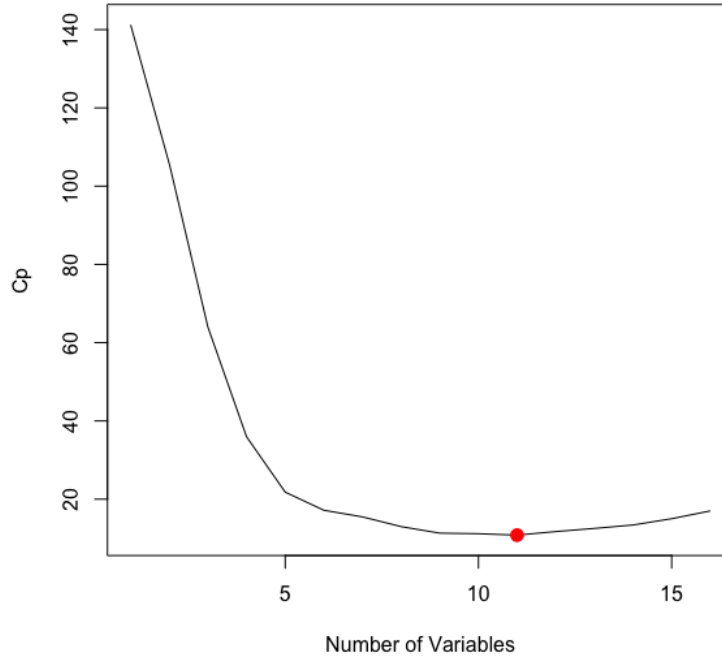
```
> coef(regfit.full,14)
```

| | | | | | |
|----------------------------|---------------|------------------------|---------------|-----------------------------|---------------|
| (Intercept) | 1.304878e+00 | minimum_investment | -4.273181e-09 | main_strategyEvent-Driven | -4.506033e-02 |
| main_strategyFund of Funds | 1.781853e-01 | main_strategyMacro | -6.149280e-02 | main_strategyRelative Value | -2.401725e-02 |
| leverage1 | -2.485841e-02 | leverage2 | -8.682929e-02 | management_fee | 6.714168e-02 |
| incentive_fee | 1.037113e-02 | redemptions | -1.000891e-04 | advance_notice | -3.547591e-04 |
| is_diversity_firmYes | -2.985147e-02 | firm_sec_registeredYes | -8.116896e-02 | geometric_return | -3.103785e+00 |

Criterion 2: Cp

Based on the Cp criterion, relative value strategy, high watermark, diversity, lockup period, and redemption period are excluded from further variable selection process.

2009 - 2014

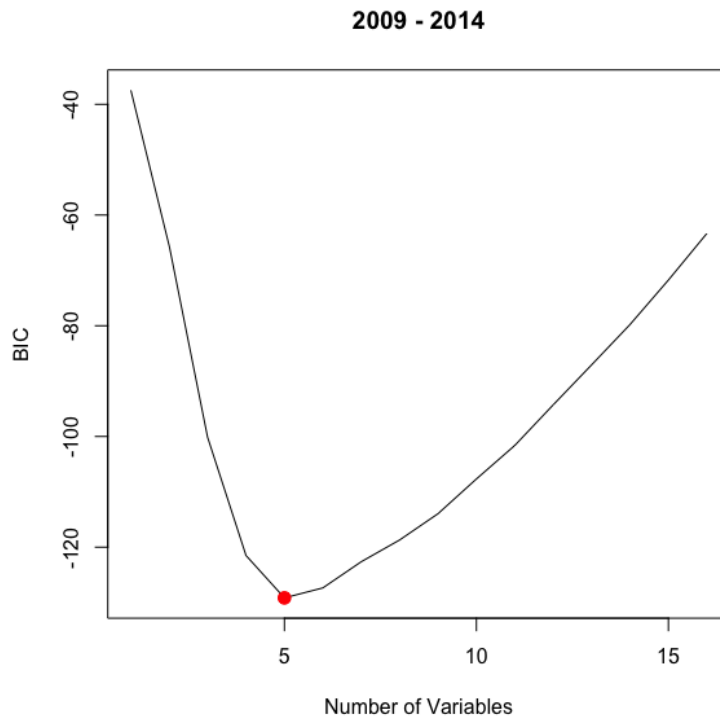


```
> coef(regfit.full,11)
```

| | | |
|----------------------------|------------------------|---------------------------|
| (Intercept) | minimum_investment | main_strategyEvent-Driven |
| 1.294552e+00 | -4.402616e-09 | -3.839979e-02 |
| main_strategyFund of Funds | main_strategyMacro | leverage1 |
| 1.840498e-01 | -5.351533e-02 | -2.464818e-02 |
| leverage2 | management_fee | incentive_fee |
| -8.944929e-02 | 6.636867e-02 | 1.044553e-02 |
| advance_notice | firm_sec_registeredYes | geometric_return |
| -5.088783e-04 | -8.101682e-02 | -3.132359e+00 |

Criterion 3: BIC

The BIC method indicates that number of variables should be included in the model is 5:



```
> coef(regfit.full,5) ## we thus use 5 variables according to BIC
      (Intercept) main_strategyFund of Funds      management_fee
      1.255941411          0.189188593          0.058979372
incentive_fee    firm_sec_registeredYes    geometric_return
      0.009823253          -0.082173495          -3.055271262
```

In addition, by reviewing the results from both modern and traditional methods, I recode main strategy into “fund and fund” and “else”. Because the model only picks out “fund of fund” strategy as different from baseline, we regroup all other strategies into one baseline strategy called “else”, while retaining “fund of fund” as a second level in the strategy variable.

Finally, I put the remaining variables into logistic regression to do inference with the output shown below. According to the result, we can see high management fee and incentive fee are negatively correlated with hedge fund’s survival probability, whereas ex-macro strategy

(labeled as “main_strategyelse” in the output), being SEC registered, and good historical performance are positively correlated with the survival probability.

```
> summary(fit)

Call:
glm(formula = dat$fund_status ~ ., family = binomial, data = dat)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6571 -1.0898 -0.8217  1.1945  2.8208

Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.244127   0.124239  -1.965   0.0494 *
main_strategyelse -0.829198   0.111613  -7.429 1.09e-13 ***
management_fee    0.255552   0.062643   4.079 4.51e-05 ***
incentive_fee     0.043976   0.005876   7.483 7.24e-14 ***
firm_sec_registeredYes -0.336954   0.065833  -5.118 3.08e-07 ***
geometric_return  -15.714611   2.343886  -6.705 2.02e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 6272.1  on 4547  degrees of freedom
Residual deviance: 6083.6  on 4542  degrees of freedom
AIC: 6095.6

Number of Fisher Scoring iterations: 4
```

As a summary, the variables that are identified as significant in two periods are not entirely the same, meaning the survival model varies from time to time, possibly due to certain change in the industry dynamics, the market situation, and the financial crisis, which I will further discuss in the next section. However, more than half of the significant variables from the pre-crisis period do remain significant in the post-crisis period, indicating there *are* some attributes, including investment strategy, incentive fee, SEC registration, historical performance, that have continuously played vital roles in a hedge funds’ survivorship.

VIII. Discussion

This section includes two parts: 1) discussion about the attributes that are identified as significant in both periods; 2) discussion about selected attributes that are identified as significant in only one period.

1) Significant Attributes in Both Periods

Main Strategy

It is necessary to note that although main strategy appear to be significant in both periods, the significant strategy for the first period is different from that for the second period. In particular, in the pre-crisis period, a hedge fund's survival probability is positively correlated with macro investment strategy, whereas in the post-crisis period, hedge fund's survival probability is positively correlated with the "else" strategies (i.e. event driven, relative value, and macro). This finding is consistent with Fung and Hsieh (2001), who find out that empirically there has not been any evidence showing there is any single investment style that can consistently outperform its peers.

Incentive Fee

Incentive fee has correlation with hedge fund's survivorship, based on results from both periods. In particular, high incentive fee is associated with low survival probability in both periods. This finding does not correspond to previous findings from scholarly research. For example, Gregoriou (2002) finds out that funds survive longer if they are large with high mean monthly returns, high performance fees, and low leverage, and this correlation holds the same for both live and defunct funds. It is reasonable that different time periods, datasets, as well as models

used in research may produce different findings. In addition, the positive correlation does not hold as before may be because 1) the hedge fund industry has become more competitive from time to time, 2) investors might not think the high fee structure is as justified as before, especially a great number of hedge fund suffered huge losses in the event of financial crisis, 3) high fee structure may incentivize manager to take excessive risk, which would potentially create detrimental impact on performance in case of unexpected market events.

Firm SEC Registered

Since information with regard to fund's asset is not completely available across the two datasets, I use "firm SEC registered" as a proxy for size of hedge fund. However, this proxy comes with limitation as the requirement to register with SEC was changed by the Dodd-Frank Act in 2012. Specifically, before the Dodd-Frank, funds can register with SEC voluntarily, regardless of the size; however, the Dodd-Frank Act requires only those hedge funds with asset more than \$100 million to register with the SEC, and all other smaller funds would have to register with their respective state regulator instead of the SEC. Since the requirement of being SEC registered has changed after the financial crisis (although before the crisis, funds that were registered with the SEC tend to be larger in size), there is limitation by using this attribute as a proxy for size.

If we treat this attribute as a proxy for size, regardless of the limitation, it appears to be positively correlated with survival probability in both periods, indicating larger-sized hedge fund tend to have higher survival probability from time to time. This conclusion echoes with Boyson's research (2002), which concludes that hedge fund size has a strong effect on survivorship and smaller funds die faster than the larger ones.

Historical Performance

Geometric average monthly return functions as a proxy for historical performance in this research. As outputs from both pre-crisis and post-crisis periods indicate, historical performance consistently has strong correlation with the survival probability (with large absolute value of coefficient and very small p-value in both periods). In the same vein, Gregoriou (2002) reaches a similar conclusion on the correlation between performance and hedge fund's survivorship. In particular, he divides all defunct hedge funds into 10 classes, depending on their investment strategies, and finds out that 1) of the 10 classes of defunct hedge funds, none had significantly positive average returns in the 12, 6, or 1 month(s) before it stopped reporting its performance; 2) all of the fund classes had greater standard deviation of returns in the previous month than 6 or 12 months earlier. These two findings both indicate that poor performance could be a major reason that funds become defunct.

2) Significant Attributes Appearing Only in One Period

Leverage

According to the logistic regression output from the first period, hedge funds that have higher leverage tend to have higher survival probability before the financial crisis. However, such correlation vanished after the crisis, which may be explained by the patterns of the US stock market. Although in both periods, the US stock market had rallied for a long period of time, it is imperative to be aware that there were two relatively large-scale stock market corrections during the second period.

From a previous research, it is found that other things equal, declines in stock prices are accompanied by larger increases in volatility than the decline in volatility that accompanies

rising stock markets (Nelson, 1991; Engle and Ng, 1993). Therefore, it is possible that high leverage fund has not benefited as much from the bull market due to high volatility embedded in the market corrections.

Diversity

According to the HFR, diversity firm is defined as a firm that is owned by either ethnic minority or woman. Based on the logistic regression output, it is important to note that this attribute shows high correlation with a fund's survivorship in the pre-crisis era, but disappeared after the financial crisis.

Surprisingly, this is an attribute that not many scholars have studied on, but insights from the industry may offer an alternative way to explain this correlation. According to a journal on affirmative investing issued by Barclays (2011), it is found that 1) the appetite for women and minority owned hedge funds appears to be growing; 2) the number of women and minority owned hedge fund launches increased from 2006 to 2008, but has retrenched in 2009 and 2010, probably due to the economic meltdown; 3) although women and minority owned hedge fund firms tend to be smaller (median \$65 million assets under management) and younger than their non-diversity peers, their performance, both in terms of absolute returns and risk-adjusted returns, is substantially stronger than the hedge fund universe at large; 4) risk aversion, while academically noted in women investors, seems to play a role in capital preservation, with women and minority owned funds outperforming their non-diversity peers in market downturns.

Management Fee

The logistic regression output indicates that low management fee is associated with high survival probability, whereas this relationship was not manifested before the crisis. A few possible reasons for this result are largely similar to the ones that have been discussed in the incentive fee section.

IX. Conclusion

In this research, I take a statistical approach to find whether some variables, individually and/or collectively, have statistically significant effects on the fund's survival probability and whether the survival probability model has changed after the financial crisis. I find out that 1) the variables that are identified as significant in two periods are not entirely the same, meaning the survival model may vary from time to time; 2) although significant variables for each period are not the same, there are some variables, including incentive fee, historical performance, and SEC registration status have tested to be consistently correlated with hedge funds' survival probability.

Works Cited

- Amin, Gaurav S., and Harry M. Kat (2003). "Welcome to the Dark Side: Hedge Fund Attrition and Survivorship Bias over the Period 1994-2001." *Journal of Alternative Investments* 6 (2003): 57-73.
- Barclays Capital (2011). Affirmative Investing: Women and Minority Owned Hedge Funds. Retrieved from http://diverseassetmanagers.org/pdf/Barclays_Capital_Solutions_Group_report_june_2011.pdf.
- Bares, Pierre-Antoine, Rajna Gibson, and Sebastien Gyger (2001). "Style consistency and survival probability in the hedge funds industry." *EFA 2001 Barcelona Meetings*.
- Boyson, Nicole (2002). "How are hedge fund manager characteristics related to performance, volatility and survival" *Unpublished working paper, Ohio State University, Fisher College of Business*.
- Brown, S. J., Goetzmann, W. N. and Park, J. (2001) 'Careers and Survival: Competition and Risk in the Hedge Fund and CTA Industry', *Journal of Finance*, 56(5), 1869–86.
- Engle, R. F., Ng, V. K., (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance* 48, 1749–1778.
- Fung, W., & Hsieh, D. A. (2000). Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases. *Journal of Financial and Quantitative analysis*, 35(03), 291-307.
- Fung, W., & Hsieh, D. A. (2001). The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial studies*, 14(2), 313-341.
- Getmansky, Mila. (2004). "The Life Cycle of Hedge Funds: Fund Flows, size and Performance." Unpublished Working Paper, MIT Laboratory for Financial Engineering.
- Gregoriou, Greg N. (2002). "Hedge Fund Survival Lifetimes." *Journal of Asset Management* 3.

- Gregoriou, G. N. and Rouah, F. (2002) 'Is Size a Factor in Hedge Fund Performance?', *Derivatives Use, Trading and Regulation*, 7(4), 301–5.
- Gregoriou G, Lhabitant F, Rouah FD (2009). The survival of exchange-listed hedge funds. *Journal of Applied Research in Accounting and Finance* 4: 2-1.
- Howell, M. J. (2001) 'Fund Age and Performance', *Journal of Alternative Investments*, 4(2), 57–60.
- Jen, Patricia, Chris Heasman, and Kit Boyatt (2001). "Alternative asset strategies: early performance in hedge fund managers." *Lazard Asset Management*.
- Kat, Harry M., and Chris Brooks (2001). "The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors." *Cass Business School Research Paper*.
- Lee, Hee Soo, and Juan Yao (2015). "Evaluating and predicting the failure probabilities of hedge funds." *Investment Analysts Journal* (2015): 1-17.
- Liang, B. (2000). Hedge funds: The living and the dead. *Journal of Financial and Quantitative analysis*, 35(03), 309-326.
- Liang, B., & Park, H. (2010). Predicting hedge fund failure: A comparison of risk measures.
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59, 347–370.
- Simon G. (2011). Does lockup increase hedge funds lifetimes? Working Paper, University of Toulouse.