JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **26**: 193–212 (2011)
Published online 7 January 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.1147

ASSESSING AND VALUING THE NONLINEAR STRUCTURE OF HEDGE FUND RETURNS

ANTONIO DIEZ DE LOS RIOS[a] AND RENÉ GARCIA[b]*
[a] *Financial Markets Department, Bank of Canada, Ottawa, Canada*
[b] *Finance, Law and Accounting Department, EDHEC Business School, Nice, France*

SUMMARY

Several studies have put forward that hedge fund returns exhibit a nonlinear relationship with equity market returns, captured either through constructed portfolios of traded options or piece-wise linear regressions. This paper provides a statistical methodology to unveil such nonlinear features with respect to returns on benchmark risk portfolios. We estimate a portfolio of options that best approximates the returns of a given hedge fund, account for this search in the statistical testing of the nonlinearity, and provide a reliable test for a positive valuation of the fund. We find that not all fund categories exhibit significant nonlinearities, and that only a few strategies provide significant value to investors. Our methodology helps identify individual funds that provide value in an otherwise poorly performing category. Copyright © 2010 John Wiley & Sons, Ltd.

Received 3 December 2007; Revised 22 June 2009

1. INTRODUCTION

Many pension funds invest in hedge funds with the hope of improving their performance. However, episodes involving long-term capital management (LTCM) in 1998 and Amaranth in 2006 have raised questions about the true nature of their risks.[1] Recent literature suggests that hedge fund returns exhibit nonlinear structures and option-like features. Fung and Hsieh (2001) analyze trend-following strategies and show that their payoffs are related to those of an investment in a lookback straddle. Mitchell and Pulvino (2001) show that returns to risk arbitrage are similar to those obtained from selling uncovered index put options. Agarwal and Naik (2004) extend these results and show that, in fact, a wide range of equity-oriented hedge fund strategies exhibit this nonlinear payoff structure.

Previous studies aiming at uncovering nonlinearities feature several shortcomings. In Agarwal and Naik (2004), linear and nonlinear exposures to risk factors are determined in a stepwise regression approach, whereby variables are added or deleted in a sequential way based on the value of the F-statistic. Such a search makes it impossible to rely on standard statistical inference. Another potential limitation is the fact that the nonlinear risk exposures are chosen a priori, by including at-the-money and out-of-the-money puts and calls on the S&P 500 index. Indeed, hedge funds do not hold the same portfolio of options since their investment strategies differ considerably. Therefore, both the number and the strike price of these options need to be determined for each fund (see Amin and Kat, 2003). Also, managers can use strategies to replicate synthetically the

* Correspondence to: René Garcia, Accounting, Law and Finance Department, EDHEC Business School, 393 Promenade des Anglais, BP 3116, Nice Cedex 3, 06202 France. E-mail: rene.garcia@edhec.edu
[1] See an excellent survey by Stulz (2007) and the recent references therein.

payoffs of options on a benchmark portfolio other than the S&P500, for which no liquid options exist.

We propose an econometric methodology that overcomes the limitations mentioned above. Our approach allows us to (i) use options on any benchmark portfolio deemed to best characterize the strategies of the fund (and not simply traded options on the S&P 500 or other liquid options), (ii) estimate whether the options that best describe the returns of a particular fund are puts or calls, or both, as well as their corresponding moneyness, (iii) assess whether the presence of the estimated nonlinearities are statistically significant, over and above the linear factors, (iv) value the performance of a fund by valuing the portfolio of options that have been found to be significant in characterizing the hedge fund returns, and (v) provide a reliable test for a positive valuation of the fund.

We base our methodology on Glosten and Jagannathan (1994) but we estimate a flexible piece-wise linear function, instead of setting it a priori, to capture the potentially nonlinear relationship between the returns of a hedge fund and those of a set of benchmark risk portfolios. Following Hasanhodzic and Lo (2007), we choose to include stocks, bonds, currencies, commodities, and credit as risk factors. Given the small sample of hedge fund return observations, we believe that this set of factors provides a reasonable trade-off between the right number of risk exposures for a typical hedge fund and a potential over-fitting of the model. Our additional contribution with respect to Glosten and Jagannathan (1994) is to propose a valid inference procedure in such a framework. Indeed, standard hypothesis tests to determine whether the coefficients that capture the positions on the options are different from zero are not applicable. To overcome this problem, we adapt a methodology proposed by Hansen (1991, 1996) and compute the critical values corresponding to the appropriate asymptotic distribution.

We apply this methodology to several indexes of hedge fund categories, such as convertible arbitrage, fixed-income arbitrage, event-driven, equity market neutral, long–short equity, global macro, and managed futures. We conduct our analysis after correcting for biases that affect reported data on hedge fund returns. We also extend previous studies by applying the methodology to individual funds within the categories to unveil the positiveness of manager-specific performance.[2] Furthermore, we account for data snooping when analyzing individual funds.

Our findings indicate that using a proper statistical methodology matters. We find that there is statistical support for rejecting linearity only for a few categories. This conclusion differs from Agarwal and Naik (2004), who find evidence of nonlinearities in most equity-related indexes. For valuation, after correcting for the backfilling and liquidation biases, only emerging markets, dedicated short bias and managed futures do not present a positive valuation. Moreover, since our methodology relies on using an option-pricing model to value the nonlinear risk exposures, we verify that it leads to similar findings to studies based on observed option portfolios when they are available. Looking at individual funds, we point out that results based only on indexes are misleading. The appearance of nonlinear features in hedge fund returns is supported statistically for only a fifth of the individual funds. Only one fund out of two provides a significant positive performance to its investors. In addition, we find that there is cross-sectional variation in the estimated moneyness across individual funds. These conclusions emphasize that both testing and disaggregation are important to draw a realistic picture of performance in the hedge fund industry.

[2] These studies limited themselves to indexes of hedge funds or considered individual funds in one particular category. Mitchell and Pulvino (2001) look at individual funds in the sole risk arbitrage category, while Fung and Hsieh (2001) study only trend-following strategies. Patton (2007), Chan *et al*. (2007) and Hasanhodzic and Lo (2007) also look at returns from individual funds.

DOI: 10.1002/jae

There are also important variations between the strategies in terms of detected nonlinearities and positive performance.

Other papers have recently proposed statistical tests to assess the risk and the performance of hedge funds. Patton (2007) investigates whether hedge funds in the market neutral category are really market neutral. Chan *et al*. (2007) develop a number of new risk measures for hedge funds, such as illiquidity risk exposure and nonlinear factor models, and apply them to individual and aggregate hedge-fund returns.

The rest of this paper is organized as follows. Section 2 describes the tests used to assess the presence of nonlinearities and to value them, as well as to control for data snooping. Section 3 describes the data and presents results for a global index, style indexes, and individual funds in several groupings of strategies. Section 4 provides results of a simulation study to assess the finite-sample properties of the linearity test. Section 5 concludes.

2. ASSESSING AND VALUING NONLINEARITIES

Our approach follows Glosten and Jagannathan (1994), who suggest approximating the payoff on a managed portfolio using payoffs for a limited number of options on a suitably chosen index portfolio and evaluating the performance of a managed portfolio by finding the value of these options.

2.1. Assessing the Nonlinearities

We start by fitting a piecewise linear function such as:

$$X_{p,t} = \beta_0 + \beta' R_{I,t} + \sum\nolimits_{i=1}^{m} \delta_j \max(R_{m,t} - k_j, 0) + \varepsilon_t \qquad t = 1, \ldots n \tag{1}$$

which can be interpreted as a regression equation of the excess return of a hedge fund ($X_{p,t}$) on a constant, the returns on a set of benchmark portfolios explaining hedge fund returns ($R_{I,t}$), and the payoffs at expiration of m call options on a diversified equity index portfolio (with return $R_{m,t}$) but with different strikes.[3] However, such an approach requires the specification of the number of options, m, and their strikes $\{k_1, \ldots, k_n\}$. In previous papers, such as Glosten and Jagannathan (1994) and Agarwal and Naik (2004), these are chosen a priori.[4] Here, we want to let the data optimally determine the values of these parameters. We will show that this extra degree of flexibility is critical to characterize the strategy followed by hedge funds and to evaluate their performance.

To determine the number of options needed to approximate well the returns of a hedge fund, we start by testing whether the linear fit ($m = 0$) provides a better approximation to the description of the data than a model with only one option ($m = 1$). If we cannot reject the hypothesis that the model is linear, we can stop there. Otherwise, we could test whether the fit of a model with two options is better than the fit with only one, and so on. In particular, let us assume

[3] Note that $R_{m,t}$ is one of the variables included in $R_{I,t}$.

[4] Glosten and Jagannathan (1994) set the knot equal to one for a one-knot estimation, as in Henriksson and Merton (1981). Agarwal and Naik (2004) do not use the same estimation strategy. They compute the returns of strategies based on options using the observed prices of calls and puts at the money and slightly out-of-the-money for the S&P 500 index.

DOI: 10.1002/jae

for the sake of simplicity that there is only one benchmark portfolio: an equity index portfolio $R_{I,t} = R_{m,t}$. Then, when $\delta = 0$, the linear model is nested in the formulation with one option, $m = 1$:

$$X_{p,t} = \beta_0 + \beta_1 R_{m,t} + \delta \ \max(R_{m,t} - k, 0) + \varepsilon_t \qquad t = 1, \ldots n \tag{2}$$

But if the model is linear, the strike of the option is not identified, meaning that any value of k will leave the R^2 of the regression unchanged. Therefore, the asymptotic distribution of the usual test statistic of the hypothesis that δ is equal to zero is not standard, which means that we cannot rely on a table of known critical values, as is usually done. As a consequence, the nonlinear pattern in hedge fund returns found in previous papers may just be a statistical artifact due to an ad hoc specification of the number of options (and their strikes), or to the use of an invalid statistical testing theory.

To account for this problem, we appeal to the general theory for econometric testing problems involving parameters that are not identified under the null hypothesis developed in Hansen (1991, 1996). In particular, it is convenient to rewrite this specification as

$$y_t = \mathbf{x}_t(k)'\mathbf{b} + \varepsilon_t \qquad t = 1, \ldots n$$

where $y_t = X_{p,t}$, $\mathbf{x}_t(k) = [1, R_{m,t}, \max(R_{m,t} - k, 0)]'$ and $\mathbf{b} = [\beta_0, \beta_1, \delta]'$.

If the strike of the option k were known a priori, we could use the usual heteroskedasticity-robust Wald statistic to test if H_0: $\delta = 0$ against H_1: $\delta \neq 0$:

$$T_n(k) = n\widehat{\mathbf{b}}(k)'R[R'\widehat{\mathbf{V}}(k)R]^{-1}R'\widehat{\mathbf{b}}(k) \tag{3}$$

where $\widehat{\mathbf{b}}(k) = \left[\sum_{t=1}^{n} \mathbf{x}_t(k)\mathbf{x}_t(k)'\right]^{-1} \left[\sum_{t=1}^{n} \mathbf{x}_t(k)y_t\right]$ and R is the vector that, applied to the vector $\mathbf{b}$, selects the parameter of interest, δ; that is, $R = (0, 0, 1)'$. The robust estimate of the covariance matrix $\widehat{\mathbf{V}}(k)$ is of the usual form $\widehat{\mathbf{M}}(k, k)^{-1}\widehat{\mathbf{K}}(k, k)\widehat{\mathbf{M}}(k, k)^{-1}$, where

$$\widehat{\mathbf{K}}(k_1, k_2) = \frac{1}{n}\sum_{t=1}^{n}[\widehat{\mathbf{s}}_t(k_1)\widehat{\mathbf{s}}_t(k_2)'],$$

$$\widehat{\mathbf{M}}(k_1, k_2) = \frac{1}{n}\sum_{t=1}^{n}[\mathbf{x}_t(k_1)\mathbf{x}_t(k_2)']$$

with $\widehat{\mathbf{s}}_t(k) = \mathbf{x}_t(k)[y_t - \mathbf{x}_t(k)'\widehat{\mathbf{b}}(k)]$ being the regression score evaluated at the sample optimal parameter estimate, $\widehat{\mathbf{b}}(k)$. The Wald statistic $T_n(k)$ will have an approximate chi-square distribution with one degree of freedom (the number of restrictions) in large samples.

The least-square estimate of k can be found sequentially through concentration. For a given value of the strike of the option k, we run an OLS regression as if k were known. We then search over the possible values of k for the one that minimizes the sum of squared errors $\widehat{\varepsilon}(k)'\widehat{\varepsilon}(k)$ to get our estimate of this parameter.[5] However, k is chosen in a data-dependent procedure and, therefore, the chi-square distribution for the Wald test statistic is invalid. Thus we follow Davies (1977,

[5] Following Hansen (1996, 1999), we restrict our search to the observed values of x_t. Moreover, since the point-wise statistics are ill behaved for extreme values of k, we further restrict the search to values of x_t lying between the τth and $(1 - \tau\text{th})$ quantiles of its distribution, with $\tau = 0.15$.

1987), who suggests computing the Wald test statistic, $T_n(k)$, for each possible value of k and then focusing on the supremum value of such a sequence; that is, $T_n = \sup_k T_n(k)$. This statistic is known as 'supWald' and it has an asymptotic distribution that is non-standard. In particular, using empirical process theory, Hansen (1996) derives the asymptotic distribution of this test under the null hypothesis and provides a simulation method to compute the various distributions. He shows that the test statistic sequence, $T_n(k)$ (e.g., the Wald test for each possible value of the strike k), converges in distribution to the following process:

$$T_n(k) \longrightarrow_d T(k),$$
$$T(k) = \mathbf{S}(k)'\mathbf{M}(k,k)^{-1}R[R'\mathbf{V}(k)R]^{-1}R'\mathbf{M}(k,k)^{-1}\mathbf{S}(k)$$

where $\mathbf{S}(k)$ denotes a mean zero Gaussian process with a covariance kernel $\mathbf{K}(k_1, k_2)$,[6] such that $\mathbf{S}_n(k) = (1/\sqrt{n})\sum_{t=1}^{n}\mathbf{s}_t(k)$ converges in distribution to $\mathbf{S}(k)$. This implies that the supWald statistic T_n converges to $T = \sup_k T(k)$, and Hansen (1996) proposes calculating the asymptotic distribution of this statistic T through simulation. In particular, let J be the number of simulations used to approximate the asymptotic distribution of the statistic.[7] Then, for $j = 1, \ldots, J$, execute the following steps:

1. Generate $\{u_{tj}\}_{t=1}^{n}$ i.i.d. $N(0,1)$ random variables.
2. Set $\mathbf{S}_n^j(k) = \dfrac{1}{\sqrt{n}}\sum_{t=1}^{n}\widehat{\mathbf{s}}_t(k)u_{tj}$.
3. Set $T_n^j(k) = \mathbf{S}_n^j(k)'\widehat{\mathbf{M}}(k,k)^{-1}R[R'\widehat{\mathbf{V}}(k)R']^{-1}R'\widehat{\mathbf{M}}(k,k)^{-1}\mathbf{S}_n^j(k)$.
4. Set $T_n^j = \max_k \; T_n^j(k)$.

This gives a random sample $\{T_n^1, \ldots, T_n^J\}$ of observations of the conditional distribution of the statistic. Finally, we can compute the percentage of these artificial observations which exceed the actual test statistic T_n to compute an 'asymptotic p-value' such as

$$\widehat{p}_n^J = \frac{1}{J}\sum_{t=1}^{n}\{T_n^j \geq T_n\}$$

and, as usual, if the value of this 'asymptotic p-value' falls below the usual 10%, 5% or 1% value, then we will reject the null hypothesis of linearity at that level.

2.2. Valuing the Nonlinearities

Our goal is to find whether a given fund provides a positive performance to investors; that is, to find the fair value of the portfolio of options that replicates the hedge fund. To do so, we appeal to a no-arbitrage argument and assume the existence of a strictly positive stochastic discount factor (SDF) that prices any traded asset.[8] The SDF is denoted by M_{t+1} and its existence implies that

[6] It means that, for any $\{k_1, k_2, \ldots k_l\}$, $\{\mathbf{S}(k_1), \mathbf{S}(k_2), \ldots \mathbf{S}(k_l)\}$ is multivariate normal with mean zero and covariances $E[\mathbf{S}(k_i)\mathbf{S}(k_j)'] = \mathbf{K}(k_i, k_j)$.

[7] Again, we follow Hansen (1996, 1999) and set $J = 2000$ in our empirical exercise.

[8] Glosten and Jagannathan (1994) develop at length the arguments about the existence of the SDF starting from the marginal rates of substitution of investors. Rather, we appeal to its existence by the assumption of no arbitrage opportunities (see Hansen and Richard, 1987).

DOI: 10.1002/jae

the net present value, V_t, of a claim to the excess return of a hedge fund, $X_{p,t+1} = R_{p,t+1} - R_f$, satisfies

$$V_t = E_t[M_{t+1}X_{pt+1}] \tag{4}$$

where $E_t[\cdot]$ denotes the expectation with respect to the information available at time t. Note that V_t is the net present value at the margin of a borrowed dollar invested in the hedge fund, conditional on the information available at time t. Since the information set available at time t may be complicated, we follow Glosten and Jagannathan (1994) and focus on the average value of V_t given by[9]

$$v = E[V_t] = E[M_{t+1}X_{pt+1}] \tag{5}$$

In particular, recall that the value of a dollar for sure received at time $t+1$ is $E[M_{t+1}] = 1/R_f$; the value of $R_{i,t+1}$ received at time $t+1$ is $E[M_{t+1}R_{i,t+1}] = 1$; also note that $E[M_{t+1}\max(R_{m,t+1} - k, 0)] = C$ is the price of a call option (with one period to expiration, and exercise price k, when the current value of the stock market index is one). A natural starting point is to assume that the (gross) return on the index portfolio R_m is log-normally distributed so that the value of the option can be computed using the Black–Scholes formula. This valuation procedure has the advantage of being simple and intuitive. Thus we can combine (2) and (4) and arrive at the following value of the fund:

$$v = \beta_0/R_f + \beta_1 + \delta_1 C \tag{6}$$

and in the case with several benchmark portfolios driving the SDF as in (1) we have that

$$v = \beta_0/R_f + \beta'\iota + \delta_1 C \tag{7}$$

where ι is a vector of ones.

Given that the price of the call option, C, depends on the level of the interest rate each period, we follow Glosten and Jagannathan (1994) to estimate a normalized version of equation (2):

$$X^*_{pt} = \beta_0 + \beta_1 R^*_{mt} + \delta\max(R^*_{mt} - k, 0) + \varepsilon_t \tag{8}$$

where the asterisk means that the respective variables have been divided by R_{ft}. With this transformation, the valuation of the projection of X_{pt}, conditional on the interest rate, is independent of the interest rate. The value of the first two terms is $\beta_0 + \beta_1$, while the value of the third term can be shown (in a Black–Scholes world) to be equal to $\delta[N(d_1) - kN(d_2)]$ with $d_1 = -\log(k)/\sigma + \sigma/2$ and $d_2 = d_1 - \sigma$, where σ denotes the standard deviation of the index returns and $N(\cdot)$ is the standard normal distribution function. Thus

$$v = \beta_0 + \beta_1 + \delta[N(d_1) - kN(d_2)] \tag{9}$$

where k is now a strike on the normalized returns on the market. With the average value of the monthly interest rate, the at-the-money strike ($1/R_f$) will be equal to 0.9969. This will be useful

[9] This simplification is appropriate in a framework where the hedge fund manager accepts a dollar from the investor at time t and returns $R_{p,t+1}$ dollars at time $t+1$, and where this process is repeated for several periods. We will therefore use the time series of returns of the hedge fund along with the returns on the index to attribute an average value to the fund. We also assume that, even if the manager's abilities change over time, the average ability is still well defined.

DOI: 10.1002/jae

for interpreting the estimated value for k (see Section 3.2). If it is greater than this value, the option will be out-of-the-money.

Once we have estimated equation (8) using the techniques described in the previous section, we can plug our estimates of β_0, β_1, and δ into (9) to find the value of the fund, and test H_0: $v = 0$ against H_1: $v > 0$. Since this is a one-sided hypothesis, we follow Hansen (1991) to define the following sequence of pointwise t-statistics:

$$t_n(k) = \frac{v[\widehat{\mathbf{b}}(k)]}{\left[\dfrac{\partial v'}{\partial \mathbf{b}} \widehat{\mathbf{V}}(k) \dfrac{\partial v}{\partial \mathbf{b}}\right]^{1/2}}$$

where $v[\widehat{\mathbf{b}}(k)]$ is the value of the fund evaluated at our estimate $\widehat{\mathbf{b}}(k)$. That is, we compute the t-statistic, $t_n(k)$, for each possible value of k to focus on the supremum value of such a sequence; that is, $t_n = \sup_k t_n(k)$. We name this statistic 'sup-t'. Again, the asymptotic distribution is non-standard, and its computation requires the use of simulation methods.[10] In finance, this test for positive value of v will answer the question whether a fund is generating a positive alpha (a measure of risk-adjusted performance). In our case, we measure risk exposure by a linear relationship with a set of risk factors plus a nonlinear option-like exposure to some of these risk factors.

2.3. Data Snooping

We have shown how to test for a significantly positive value of a given hedge fund, but in practice we have a universe of funds and our goal is to identify all funds providing positive performance. To fulfill this objective, we could be tempted to select all funds for which the estimated p-value falls below the usual confidence level of, say, 5%. However, such an approach would entail testing multiple hypotheses at the same time. That is, we would be testing simultaneously that each and every fund provides zero value at the same time:

$$H_{0,i} : v_i = 0 \text{ vs } H_{1,i} : v_i > 0 \qquad \forall i = 1, \ldots, S$$

where S is the total number of funds. This is a classical example of data snooping, and failing to account for such multiple hypothesis testing might result in identifying funds that do not provide positive value to investors (see Romano *et al.*, 2008). We face a similar problem when trying to identify those funds presenting nonlinear features.

In practice, one usually deals with data snooping by controlling (asymptotically) the so-called family-wise error rate (FWE): the probability of making one or more false rejections (the probability of picking a fund that does not provide positive value to the investor). However, this criterion can be too strict when the number of hypotheses under consideration is very large, as in our case with the number of hedge funds. Instead, we use a less stringent criterion and control instead the probability of making m or more false rejections (for some integer m greater than or equal to one).[11] This approach is known as the m-FWE.

[10] Details on the derivation of such an asymptotic distribution can be found in Hansen (1991). The algorithm is similar to the procedure described for the linearity test.
[11] We choose m to be equal to 5% of the number of funds under study.

DOI: 10.1002/jae

A simple way to control the m-FWE is to use the approach known as 'generalized Holm method' introduced by Hommel and Hoffman (1988), since its implementation requires only the p-values of the individual tests.[12] First, we order the individual p-values (from either the test on the existence of nonlinearities or on the value of the fund) from smallest to largest: $\widehat{p}^J_{n,1} \leq \widehat{p}^J_{n,2} \leq \ldots \leq \widehat{p}^J_{n,S}$ with their corresponding null hypotheses labeled accordingly: $H_{0,1}, H_{0,2}, \ldots, H_{0,S}$. Then $H_{0,s}$ is rejected (e.g., a fund is picked as provided positive value) at level α if $\widehat{p}^J_{n,i} \leq \alpha^{(i)}$ for $i = 1, \ldots, S$, where

$$\alpha^{(i)} = \begin{cases} \frac{m\alpha}{S} & \text{for} \quad i \leq m \\ \frac{m\alpha}{S+k-i} & \text{for} \quad i > m \end{cases}$$

3. EMPIRICAL RESULTS

3.1. Data Description and Construction of Hedge Fund Indexes

Our hedge fund returns are computed from the TASS/Lipper database, which provides monthly returns and net asset value data on 4606 funds beginning in February 1977. For building hedge fund indexes, we start our sample in January 1996 and end it in March 2004.[13] For individual funds, we use all funds for which at least 60 observations are available. Individual funds are classified into 10 categories based on their reported investment strategies: (1) convertible arbitrage; (2) fixed-income arbitrage; (3) event-driven; (4) equity market neutral; (5) long–short equity; (6) global macro; (7) emerging markets; (8) dedicated short bias; (9) managed futures; (10) funds of funds. Detailed descriptions of investment strategies followed by hedge funds in these categories can be found in Lhabitant (2004).

The database includes, for each fund, an entry date, an exit date (if any), a date for first reporting, reasons for the fund death (if necessary), and lock-up periods. We use this information to correct two well-known biases associated with hedge fund data. The first is a backfilling or instant-history bias, whereby the database backfills the historical return data of a fund before its entry into the database. The hypothesis is that a manager will report to the database vendor only after obtaining a good track record of returns over the first periods of the hedge fund life. Consequently, we eliminate all data that precede the fund entry date in the database. A similar approach is used by Fung and Hsieh (2001). Second, we correct for the so-called liquidation bias. Many funds disappear from the database during the sample period for various reasons that may not have the same consequences in terms of monetary loss for the investor. An issue of concern is when an under-performing fund ceases to report in order to hide bad results and avoid a massive withdrawal from investors. We correct the returns for this liquidation bias by applying an extra zero return when the indicated reasons are suggestive of such a behavior.[14] This number is consistent with

[12] Alternatively, we could have used bootstrap methods such as White's (2000) 'bootstrap reality check' or Romano and Wolf's (2005) 'stepwise multiple testing' to increase power. However, to the best of our knowledge, the conditions for the validity of the bootstrap have not been verified in our setup.

[13] We choose to start in 1996 to have a reasonable representation for all categories of funds. Also, the TASS database does not give any information on exited funds prior to 1994.

[14] In particular, we apply this extra zero return when the indicated reasons for not reporting are fund liquidation, fund not reporting to TASS, managers not answering requests, and other. In all other cases, particularly mergers and dormant funds, we do not apply any loss.

DOI: 10.1002/jae

the findings of Ackermann *et al.* (1999), who find a negligible impact of liquidation and time biases.[15]

For the risk factors, following Hasanhodzic and Lo (2007), we choose five factors that provide a reasonable set of risk exposures for typical hedge funds. We include the returns on (i) the CRSP value-weighted NYSE, AMEX, and NASDAQ combined index as a stock market measure; (ii) an equally weighted portfolio of British, German and Japanese 1-month eurocurrency deposits to capture any exposure to an exchange rate (FX) factor; (iii) the Lehman US Corporate AA Intermediate Bond Index to capture bond market risk; (iv) the Lehman US[16] Corporate BAA Intermediate Bond Index in excess of the return on the Lehman US Treasury index to capture a credit risk factor; (v) the Goldman Sachs commodity index. Summary statistics for these factors are reported in panel (a) of Table I.

We build equally weighted (EW) hedge fund indexes following the methodology used for the Hedge Fund Research indexes. This approach gives relatively more weight to small hedge funds. We construct these indexes starting from the individual funds and correcting the two above-mentioned biases (backfilling and liquidation). Panel (b) of Table I reports summary statistics for the returns of these indexes in excess of the 30-day Treasury Bill yield obtained from the CRSP RISKFREE files. Mean values are all positive (except C8) but vary substantially across categories. Standard deviations are also widespread, with the short bias category exhibiting the largest volatility and the equity market neutral the lowest. Skewness alternates between negative and positive values for the various strategies. It is close to zero for the global index, where averaging tends to make the returns distribution more symmetric. Similarly, excess kurtosis is much higher in the category indexes than in the global index.

3.2. Assessing Nonlinearities and Performance in Hedge Fund Indexes

We estimate the one-option specification in (8) for the global hedge fund indexes and each of the category indexes.[17] We report four sets of results for each index: (i) the estimated values of the coefficients—respectively, the intercept (β_0), the coefficients of the five risk factors (β_1 to β_5), the coefficient on the option on the equity index (δ), and the strike (k); (ii) the R^2 and adjusted R^2 of the regressions; (iii) the test results for the presence of the nonlinearity; (iv) the first-order autocorrelation of the residuals and (v) the test results for the fund valuation.[18]

---

[15] This extra return is applied in the month following the month where the fund stopped reporting. This could be somewhat optimistic given our ignorance regarding the true loss. For example, Long-Term Capital Management lost 92% of its capital from October 1997 to October 1998 and did not report to databases (see Posthuma and van der Sluis, 2003). Malkiel and Saha (2005) estimate the survivorship bias by comparing indexes of live and defunct funds and find higher estimates than other studies. Our approach appears therefore more conservative. In an appendix available from the authors, we conduct a thorough sensitivity analysis of this liquidation bias correction. Overall, the main conclusions of the paper are not affected, but the average performance decreases with the severity of the correction.

[16] Since the credit risk factor is captured using an excess return, the value of the fund for the five-factor model is equal to $v = \beta_0 + \beta_1 + \beta_2 + \beta_3 + \beta_5 + \delta[N(d_1) - kN(d_2)]$.

[17] For space considerations, we do not report results from estimating two options on the equity index and one and more options on the other risk factors. Results, available from the authors, tend to reject more elaborate nonlinear structures, except perhaps for global macro strategies where two options suggestive of a bull spread (long and short positions in two call or put options with different exercise prices) were found significant. Option features were also found with respect to the currency factor for some strategies.

[18] For space considerations, we did not report the standard errors on the coefficients and focused rather on the overall fit of the regression and the test results for the hypotheses of interest about linearity and positive performance. Another reason for not reporting the standard errors is that we cannot use the usual critical values of a Student t-test to assess significance on the coefficients related to the options as explained in section 2.1.

DOI: 10.1002/jae

Table I. Summary statistics

Panel (a) Factors

	Mean	Median	SD	Skew.	Kurt.	Min.	Max.
1-month interest rate	3.775	4.506	0.492	−0.650	−1.166	0.794	6.234
CRSP return	10.358	19.604	17.343	−0.676	0.218	−189.157	100.709
FX return	3.583	1.552	7.409	0.298	0.029	−63.749	68.691
Bond return	0.350	1.150	3.775	−0.578	1.297	−48.903	29.285
Credit return	0.297	0.322	3.130	−0.075	3.363	−33.232	38.524
Commodity return	6.032	3.659	20.500	0.167	0.398	−199.404	211.970

Note: This table shows the means, medians, standard deviations (SD), skewness (Skew,), kurtosis (Kurt.), and minimum (Min.) and maximum (Max.) of (annualized) returns for the 1-month interest rate and the set of six factors during January 1996 to March 2004 (99 observations).

Panel (b) Indexes by categories

	Mean	Median	SD	Skew.	Kurt.	Min.	Max.
Hedge Fund Global Index	8.500	7.889	6.889	−0.233	2.441	−87.552	81.579
C1 Convertible arbitrage	10.642	11.785	4.211	−1.384	7.190	−66.418	44.404
C2 Fixed-income arbitrage	6.851	10.063	4.213	−3.157	15.657	−80.968	34.321
C3 Event-driven	9.391	10.563	5.533	−1.572	7.464	−93.348	51.167
C4 Equity market neutral	7.549	6.902	2.919	0.729	2.427	−15.590	50.059
C5 Long–short equity hedge	11.940	11.850	10.431	0.193	1.881	−107.390	129.230
C6 Global macro	6.287	6.316	6.260	0.309	0.367	−53.149	62.889
C7 Emerging markets	11.452	16.238	17.180	−1.254	4.580	−272.283	155.101
C8 Dedicated short bias	−0.026	−10.351	20.901	0.463	0.809	−186.120	252.028
C9 Managed futures	5.853	0.753	9.072	0.174	−0.011	−70.607	84.754
C10 Funds of funds	6.693	7.214	6.209	−0.194	3.081	−80.050	72.484

Note: This table shows the means, medians, standard deviations (SD), skewness (Skew,), kurtosis (Kurt.), and minimum (Min.) and maximum (Max.) of (annualized) returns for equally weighted portfolio indexes corrected by backfilling and liquidation bias, for each of the categories during January 1996 to March 2004 (99 observations).

Indexes for Fund Categories

Table II reports the estimated values for the coefficients of equation (8) for each category. Particularly, estimated values for the nonlinear component δ vary across categories. While they are mostly negative, their magnitudes differ between strategies. For managed futures and dedicated short bias, the coefficient δ is positive and β_1 is negative, while it is the opposite for the other categories. There is also cross-sectional variability among the estimated coefficients for the other risk factors. For example, managed futures have a stronger positive exposure to the FX factor than the equity market neutral category. The coefficient on the FX factor is negative for the other categories. The exposure to the bond factor is positive for all categories except long–short equity. Equity market-neutral, long–short equity and managed futures are exposed negatively to the credit risk factor. The coefficient capturing the exposure to the commodity factor tends to be small. Finally, the estimated value of the strike tends to be greater than the at-the-money strike benchmark (0.9969, when k is set to one and normalized by the monthly average of $R_{f,t}$). This means that the call option is out-of-the-money. Notable exceptions are convertible arbitrage, event-driven, and managed futures where the call option is in-the-money. The call option in the emerging markets category is near-the-money.

We also compare the increase in the goodness-of-fit obtained by using a nonlinear factor. For example, the (adjusted) R^2 is virtually unaltered for market-neutral, long–short equity, emerging

DOI: 10.1002/jae

Table II. Piecewise linear fit: indexes by category

	HF	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
Coefficients											
Constant β_0	−0.457	−0.291	−0.134	−0.360	−0.122	−0.502	−0.718	−0.686	0.823	−0.586	−0.475
CRSP β_1	−0.214	0.260	0.022	0.400	0.079	0.529	0.101	0.836	−1.041	−0.375	0.254
FX β_2	−0.082	−0.073	−0.135	−0.108	0.025	−0.003	−0.160	−0.191	−0.155	0.046	−0.129
Bond β_3	0.176	0.108	0.248	0.068	0.013	−0.098	0.768	−0.010	0.446	0.818	0.306
Credit β_4	0.180	0.344	0.462	0.270	−0.163	−0.002	0.268	0.688	0.244	−0.029	0.127
Commodity β_5	0.050	0.012	0.003	0.020	0.009	0.080	0.015	0.065	−0.071	0.081	0.047
δ	0.318	−0.224	−0.316	−0.290	−0.107	−0.199	−0.382	−0.461	0.224	0.442	−0.237
k	1.033	0.956	1.045	0.956	1.009	1.027	1.045	0.990	1.035	0.962	1.033
Adjusted R^2											
Linear	0.56	0.29	0.14	0.43	−0.01	0.67	0.20	0.44	0.67	0.17	0.42
Option	0.57	0.33	0.17	0.47	0.01	0.68	0.22	0.45	0.67	0.21	0.43
H_0: $\delta = 0$ (*p-values*)											
Wald $k = 1$	0.23	0.50	0.05	0.08	0.05	0.45	0.37	0.19	0.97	0.12	0.14
supWald	0.37	0.19	0.08	0.05	0.13	0.58	0.18	0.38	0.86	0.01	0.32
First-order autocorrelation of the residuals											
Linear	0.27	0.37	0.49	0.16	−0.01	0.12	0.15	0.35	−0.04	−0.13	0.34
Option	0.27	0.38	0.42	0.10	−0.02	0.14	0.11	0.36	−0.03	−0.12	0.32

Note: This table shows the results of the following piecewise linear fit for the equally weighted index for the different categories during January 1996 to March 2004 (99 observations): $X^*_{p,t+1} = \beta_0 + \beta' R^*_{I,t+1} + \delta \max(R^*_{m,t+1} - k, 0) + \varepsilon_{t+1}$ contains the CRSP, FX, bond, credit and commodity risk factors. HF, global hedge fund index; C1, convertible arbitrage; C2, fixed-income arbitrage; C3, event-driven; C4, equity market neutral; C5, long–short equity hedge; C6, global macro; C7, emerging markets; C8, dedicated short bias; C9, managed futures; C10, fund of funds.

markets, dedicated short bias and funds of funds. The R^2 for the other categories increases when including the nonlinear term, although the magnitude of this increase varies across categories. For example, the R^2 of managed futures increases from 0.17 to 0.21. Still, we need to know whether such an increase is statistically significant. This is precisely where the proposed testing methodology is useful. The linearity tests tell us that there is support for rejecting linearity for event-driven and managed futures at the 5% level, and fixed-income arbitrage at the 10% level. In other words, the increase in R^2 is statistically significant only for these three categories.

The level of serial correlation is high in certain fund categories such as convertible arbitrage, fixed-income arbitrage and emerging markets. However, the introduction of an option on the equity index tends to reduce the autocorrelation in some of these categories. For example, the autocorrelation in the returns of the fixed-income category falls from 0.49 with five factors to 0.42 after the introduction of the option.[19]

To interpret more easily the estimated coefficients, we show in Figure 1 several graphs that illustrate different shapes of nonlinear strategies followed by convertible arbitrage, fixed-income arbitrage, and managed futures funds.[20] Let us comment briefly on the particular strategies involved in these categories and the shapes found for the nonlinear features. A typical strategy in convertible

[19] This suggests that nonlinearities may have to be accounted for to assess actual illiquidity exposure.

[20] Test results indicate that linearity is rejected for fixed-income arbitrage and managed futures. Although linearity is not rejected at conventional levels for convertible arbitrage, we include a graph for this category because it illustrates a common strategy for hedge funds and provides the strongest reduction in p-value with respect to setting a fixed strike of one a priori (0.19 and 0.50, respectively).

DOI: 10.1002/jae

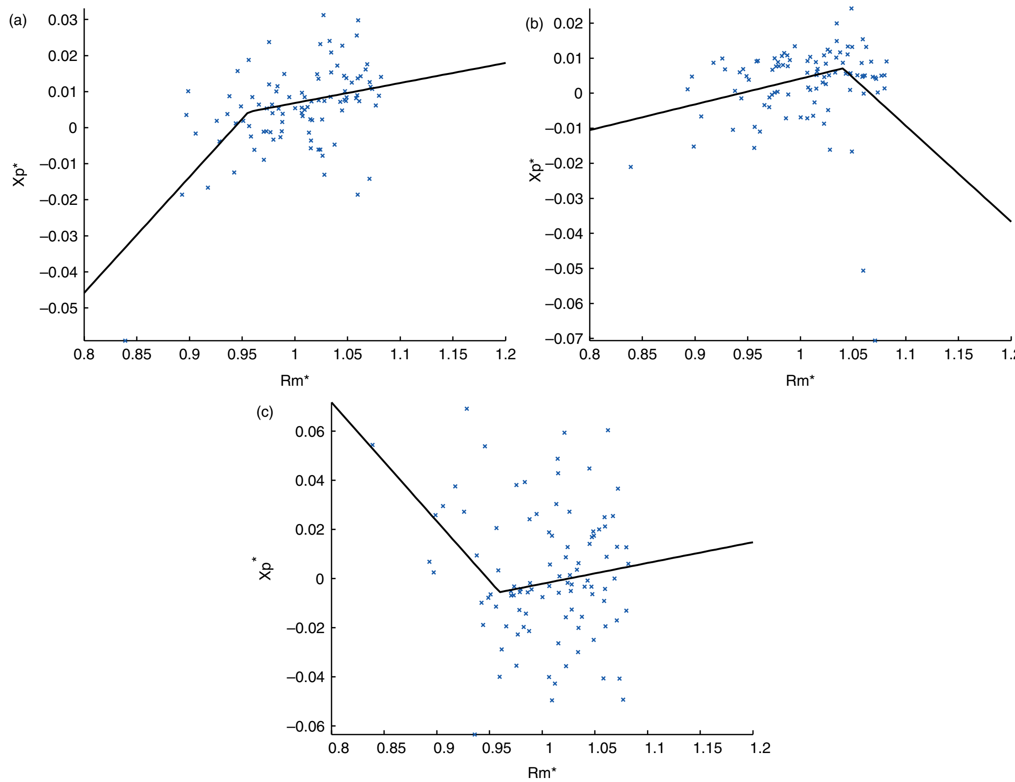

Figure 1. Piecewise linear fit. These piecewise linear fits are based on the estimates of the one-factor model in equation (8). (a) C1: convertible arbitrage. (b) C2: fixed-income arbitrage. (c) C9: managed futures. This figure is available in color online at wileyonlinelibrary.com/journal/jae

arbitrage is to be long in the convertible bond and short in the common stock of the same company. Profits are generated from both positions. The principal is usually protected from market fluctuations. The corresponding graph in Figure 1 is suggestive of a short position in a put option, which means that these strategies lose money when the equity index incurs a large fall. In most situations, however, the fund collects a small premium on the price discrepancy. On the other hand, the shape of the nonlinear feature for fixed-income arbitrage resembles that of an inverted straddle.[21] Funds in this category exploit price anomalies between related interest rate securities. They buy undervalued securities and sell short overvalued ones. The corresponding graph in Figure 1 suggests that this strategy makes money when the stock market is calm, a period during which the two securities may revert to their fundamental value. On the other hand, a large shock like the Russian crisis creates large losses. Finally, the managed futures profile in Figure 1 is more illustrative of a straddle. This result is intuitive because funds in this category tend to be trend followers. That is, they buy in an up market and sell (or even take a short position) in a down market. Therefore, large movements up or down are profitable. All these results are consistent with

[21] A straddle is an option-based strategy whereby an investor holds a position in both a call and put with the same strike price and expiration date. With an inverted straddle, the investor sells both a call and put with the same strike price and expiration date.

DOI: 10.1002/jae

Table III. Valuation: indexes by category

	α	α			Agarwal and Naik's OTM put factor
		$\sigma = 5\%$	$\sigma = 15\%$	$\sigma = 25\%$	
Hedge Fund Global Index	3.306	5.064	3.606	1.071	3.461
	[0.014]	[0.006]	[0.007]	[0.119]	[0.016]
C1 Convertible arbitrage	6.401	8.284	7.397	5.189	6.351
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]
C2 Fixed-income arbitrage	3.836	5.301	3.933	0.634	3.843
	[0.002]	[0.000]	[0.000]	[0.035]	[0.002]
C3 Event-driven	4.399	6.831	5.685	2.834	4.272
	[0.002]	[0.000]	[0.000]	[0.138]	[0.005]
C4 Equity market neutral	3.576	5.368	3.978	2.510	3.665
	[0.000]	[0.000]	[0.000]	[0.002]	[0.000]
C5 Long–short equity hedge	4.436	6.494	4.822	2.334	5.001
	[0.007]	[0.020]	[0.010]	[0.050]	[0.005]
C6 Global macro	4.615	6.386	4.733	0.746	4.698
	[0.010]	[0.003]	[0.009]	[0.029]	[0.010]
C7 Emerging markets	3.339	11.442	5.582	−0.651	2.868
	[0.235]	[0.113]	[0.173]	[0.586]	[0.276]
C8 Dedicated short bias	4.361	2.632	4.080	6.698	4.555
	[0.139]	[0.292]	[0.163]	[0.140]	[0.121]
C9 Managed futures	5.218	0.900	3.147	7.851	5.529
	[0.034]	[0.086]	[0.055]	[0.008]	[0.026]
C10 Funds of funds	2.406	4.356	2.738	−0.075	2.435
	[0.064]	[0.020]	[0.032]	[0.324]	[0.074]

Note: This table shows the value of the fund when estimating the five-factors model for different levels of annual volatility of the stock market σ. p-values for the hypothesis that the value of the fund is equal to zero, H_0: $v = 0$, against the alternative hypothesis that the value of the fund is positive, H_a: $v > 0$, are presented in brackets. The last column presents the valuation of the fund when using Agarwal and Naik's (2004) OTM put factor.

the findings of Agarwal and Naik (2004), but our method allows us to provide clear illustrations of the underlying strategies.

The ultimate test for investors remains a positive value. In Table III, we report the alpha (risk-adjusted performance) corresponding to a linear projection on the five risk factors, as well as the valuation after accounting for the option feature at various levels of volatility. We also report p-values in brackets to test whether the performance is significantly different from zero. For an annual volatility of the stock market return equal to 15%, all categories exhibit a positive performance, but for emerging markets, dedicated short bias and managed futures the value is not significantly different from zero.

Robustness to Volatility and Option Valuation

In the preceding section, we relied on the Black–Scholes formula with a constant average volatility over the sample to value the option-like nonlinear features. Thus, to show how fund value varies with the level of volatility, Table III includes two additional values of the volatility, namely 5% and 25%, when reporting tests for positive valuation. A high volatility may transform a positive performance into a negative one. For example, once volatility is at 25%, the number of categories presenting a positive valuation gets reduced. In this case, we only find a statistically positive value for convertible arbitrage, fixed-income arbitrage, long–short, global macro and managed futures. The value of the emerging markets category even gets negative. Also, it is interesting to note that performance is negatively correlated with the level of volatility of the stock market. The only exceptions to this rule are short bias and managed futures (remember the straddle-like pattern found for the latter category).

Of course, volatility varies over time and so will the value of the funds. To control for time-varying volatility, we use Agarwal and Naik's (2004) OTM put factor to obtain a measure of the overall value of a fund that does not depend on the Black–Scholes formula. In particular, we start by running a regression of the normalized hedge fund returns on the normalized benchmark indexes and the OTM put option factors:[22]

$$X^*_{pt} = \beta_0 + \beta' R^*_{It} + \beta_p R^*_{put,t} + \varepsilon_t \tag{10}$$

where the asterisk indicates that the returns on the indexes have been normalized by R_{ft}, as in equation (8); R_{It} denotes the returns on the five risk factors; and $R_{call,t}$ and $R_{put,t}$ denote the returns on the OTM call and put option factors, respectively. Note that equation (10) implies that the performance of a fund is given by $v^{AN} = \beta_0 + \beta'\imath + \beta_p$. In the last column of Table III we report the valuation of the fund using these option-based factors as well as the p-value that corresponds to the hypothesis that the value of the fund is zero. An important finding is that we arrive at the same conclusions regarding positive or negative valuation as with our methodology for volatilities between 15% and 25%. In terms of magnitude, the results are the closest for a volatility of the market equal to 20%. Therefore using the Black–Scholes model to value option-like features does not lead to different conclusions given an appropriate volatility level. Since our procedure allows valuation of options on any benchmark factor instead of relying on liquid markets, this check provides a comforting reassurance for using our methodology.

3.3. Assessing Nonlinearities and Performance in Individual Funds

Data on individual funds help unveil the reality behind indexes. Aggregation could potentially create either a smoothing effect, which will mask existing nonlinearities for each individual fund, or, at the opposite, create spurious nonlinear structures that are not present in individual funds. In this section, all funds, alive or dead, since the inception of the database in 1977, will be included in the analysis, as long as a fund has been in existence for 60 months.[23] In this section, we only correct for the liquidation bias in order to maximize the number of funds included and, therefore, we do introduce an extra zero return when a fund stops reporting for the reasons mentioned above. On the contrary, eliminating backfilled returns would have sensibly reduced the number of funds. This will undoubtedly bias performance upwards for some funds.

All Funds, Live Funds, Graveyard Funds

In the first column of Table IV, we first look at the whole universe of funds in the database since its introduction. Overall, 1868 funds have 60 observations or more. Recall that the total number

[22] The results in Agarwal and Naik (2004) suggest that only the OTM put option factor plays a role in explaining hedge fund returns (see, for example, Agarwal *et al*., 2007, for a similar argument).

[23] We consider that this is a minimum number of observations required to conduct the type of linearity and value tests we have described in Section 2.

DOI: 10.1002/jae

of funds in the database is 4606. This is more indicative of the very large number of entries in the past few years, especially for funds of hedge funds, than a large number of exits before 60 months of operation. The rate of growth in the number of funds has averaged 18% over the past 10 years and has accelerated considerably in the last 2–3 years, especially for funds of funds.

In panel (a) of Table IV we report the cross-sectional distribution of the linearity test. Panel (a) is based on the p-values for the supWald (with heteroskedasticity correction) linearity test; that is, the null hypothesis that the return of the fund has a linear relationship with the return on the market portfolio (H_0: $\delta = 0$ against H_1: $\delta \neq 0$). We include the five risk factors described in the previous sections. To summarize the test results, we report the maximum, minimum and average p-values but, more interestingly, the number (and percentage) of funds for which the p-value is less than 1%, between 1% and 5%, between 5% and 10%, and above 10%. The first important result is that we reject linearity for about one-fifth of the funds. This shows that simply relying on the global indexes may be misleading, but also that the nonlinear feature is not a statistical reality for many funds. If we further correct for any data-snooping bias, we only find 52 funds for which we could soundly reject the linear pattern. Panel (b) presents the cross-sectional distributions for performance. At an average volatility of 15%, only one out of every two funds provides a significant positive value to its investors. Even when we correct for data snooping, we still find that around 40% of funds provide positive value. Again, looking only at the indexes would have been misleading. When we average, the very good performers increase the mean value of the index. We have also computed the cross-sectional distribution of the estimated moneyness parameter. The average estimated k is close to one (1.0013), with a standard deviation of 0.0384. This cross-sectional variation emphasizes that estimation of the moneyness that best approximates the returns of a fund is important to draw a realistic picture of the hedge fund industry.

Columns two and three of Table IV report the linearity and performance test results separately for live and graveyard funds. In panel (a), the picture for the linearity test is similar for both live and graveyard funds. Live funds exhibit somewhat less significant nonlinearities than the aggregate, and corresponding graveyard funds somewhat more. This is not the case, of course, for performance reported in panel (b). For about three-quarters of the live funds, performance is significantly positive at 10%. This percentage falls to 20% for the graveyard funds, which means that disappearance from the database is usually associated with bad outcomes.

Arbitrage-Based Strategies

In the TASS/Lipper database, arbitrage strategies are grouped into three categories: convertible arbitrage, fixed-income arbitrage, and event-driven. In the event-driven category, the arbitrage is conducted whenever firms are merged, liquidated, bankrupt, or reorganized. Overall, the database contains 335 funds in these three categories. Event-driven funds represent about half the number of funds in this group. Results in the fourth column of Table IV, panel (a), show that close to 35% of the funds exhibit a significant nonlinearity with respect to the market return. In terms of performance (Table IV, panel (b)), a significant positive value is found for close to 85% of the funds. We still select 243 funds (out of 335) when we take data snooping seriously.

These results confirm the conclusions of Mitchell and Pulvino (2001) in their thorough study of risk arbitrage. In particular, they suggest that three parameters (two thresholds on low and high returns together with a parameter for normal returns), estimated with a piecewise linear regression, should be used in evaluating return series generated by risk arbitrage hedge funds. However, their approach does not account for the fact that the threshold is determined endogenously. We have

DOI: 10.1002/jae

Table IV. Results with individual funds: comparison

Panel (a) Cross-sectional distribution of linearity tests (p-values)

	All funds	Live funds	Graveyard funds	Arbitrage funds	Mkt. ntrl/LS funds	Directional funds
Max. p-value	1.000	1.000	0.997	0.996	1.000	0.997
Average p-value	0.419	0.439	0.379	0.313	0.497	0.400
Min. p-value	0.000	0.000	0.000	0.000	0.000	0.000
No. and % of funds						
p-value >10%	1497	1010	487	218	533	375
	80.14%	82.11%	76.33%	65.08%	88.10%	80.13%
5% < p-value <10%	140	91	49	37	31	40
	7.49%	7.40%	7.68%	11.05%	5.12%	8.55%
1% < p-value <5%	138	81	57	46	28	27
	7.39%	6.59%	8.93%	13.73%	4.63%	5.77%
p-value <1%	93	48	45	34	13	26
	4.98%	3.90%	7.05%	10.15%	2.15%	5.56%
Total no. of funds	1868	1230	638	335	605	468
No. of significant funds	52	32	44	42	11	26

Note: This table shows the value of the cross-sectional distribution of the p-values for the supWald linearity tests of those funds with at least 60 observations. Each p-value is for the hypothesis that the return of the fund has a linear relationship with the return of the market portfolio (H_0: $\delta = 0$ against H_1: $\delta \neq 0$).

Panel (b) Cross-sectional distribution of hedge fund performance

	All funds	Live funds	Graveyard funds	Arbitrage funds	Mkt. ntrl/LS funds	Directional funds
Average value	7.711	8.838	7.490	7.927	9.043	8.933
Standard deviation value	10.463	7.980	11.614	9.789	8.819	11.975
Min. value	−103.940	−22.895	−103.940	−103.940	−22.927	−103.940
First quartile (25%)	3.061	4.413	2.306	3.976	4.272	3.387
Second quartile (50%)	6.918	7.495	6.786	7.356	7.730	7.505
Third quartile (75%)	11.391	11.644	11.599	10.959	12.440	13.518
Max. value	107.200	60.840	107.200	62.291	62.291	107.200
Max. p-value	1.000	0.999	1.000	1.000	0.998	1.000
Average p-value	0.162	0.094	0.293	0.083	0.161	0.238
Min. p-value	0.000	0.000	0.000	0.000	0.000	0.000
No. and % of funds						
p-value >10%	645	288	357	50	224	229
	34.53%	23.42%	55.96%	14.93%	37.03%	48.93%
5% < p-value <10%	132	90	42	11	53	49
	7.07%	7.32%	6.58%	3.28%	8.76%	10.47%
1% < p-value <5%	264	187	77	26	121	61
	14.13%	15.20%	12.07%	7.76%	20.00%	13.03%
p-value <1%	827	665	162	248	207	129
	44.27%	54.07%	25.39%	74.03%	34.22%	27.56%
Total no. of funds	1868	120	638	335	605	468
No. of significant funds	710	597	125	243	163	100

Note: This table shows the cross-sectional distribution of the (annualized) performance (%) and the cross-sectional distribution of the p-values for the hypothesis that the value of the fund is equal to zero (H_0: $v = 0$ against H_1: $v > 0$) for those funds with at least 60 observations. The value of the fund is computed under the assumption that the annual volatility of the return on the market index is $\sigma = 15\%$.

provided a statistical procedure that accounts for this search and for uncertainty in the estimation of the parameters in question.

Equity Market-Neutral and Long–Short Strategies

In this grouping, the long–short funds are by far the most numerous. The latter category represents a large percentage (close to 30%) of the hedge fund industry. This strategy corresponds to the original strategy followed by Albert Winslow Jones in 1949. Overall, we find very little evidence of nonlinearities in this category. Results in the fifth column of Table IV, panel (a), show that between 10% and 15% of the funds exhibit a significant nonlinear relationship with the return of the CRSP index. This number is even lower if we correct for data snooping: around 2%. This confirms the results based on the index where the p-values associated with the two categories included in this grouping are among the highest. It also confirms the analysis of Agarwal and Naik (2004), who do not find a significant relationship with their option return index. In terms of performance (Table IV, panel (b)), the evidence is mixed. One fund out of two provides a significant positive value to investors, but once we correct for data snooping we can only select one fund out of four.

Directional Strategies

This grouping includes three categories: global macro, emerging markets, and managed futures. All strategies associated with these funds involve some kind of bet on market direction. This bet can be based on economic fundamentals or on some technical analysis. A common strategy, studied in detail in Fung and Hsieh (2001), is the trend-following strategy. It consists in buying in an up market and selling in a down market. This strategy is characteristic of managed futures funds, which make up half of the 468 funds included in our grouping. In the sixth column of Table IV, panel (a), we report that the nonlinear feature is present at the 10% level in about 20% of the funds. This, therefore, confirms the evidence in the index for emerging markets and managed futures. For performance (Table IV, panel (b)), the percentage of p-values below 10% for positive value is about 60%. The distribution of values indicates the widespread nature of value in this category, with extreme negative minimum and maximum values.

4. MONTE CARLO EXPERIMENTS

We now briefly describe the experimental setting and the results of a Monte Carlo simulation study to assess the finite-sample properties of the linearity tests. The parameter design is based on the estimated parameters of a one-factor model for category 9 (managed futures). In particular, we investigate two sample sizes: $T = 100$ (which is roughly our sample size for data on indexes) and $T = 60$ (the minimum number of observations that we require to include an individual fund in our study), and we conduct tests of size 10%, 5%, and 1%. However, we will refer only to those of size 5%, since no qualitative differences are observed. The number of replications is set to $N = 2000$.[24]

The market return R^*_{mt} and the error term are generated from two independent Gaussian distributions. The hedge fund return is generated according to the piecewise linear function in

[24] The number of internal simulations to compute the p-values is set to $J = 1000$, and the search is restricted to the values of x_t lying between the τth and $(1 - \tau\text{th})$ quantiles, with $\tau = 0.15$.

DOI: 10.1002/jae

equation (8). The values of β_0, β_1, k, and σ^2 are set to their corresponding estimates, while the variance of R^*_{mt} is set to the unconditional variance of this variable during the period January 1996 to March 2004. In order to assess the finite-sample size of these tests, we start by setting $\delta = 0$ (null hypothesis of linearity).

Results for this simulation study are reported in Table V. Here, we compare test statistics using two different covariance matrices: (i) heteroskedasticity-consistent LM, and (ii) heteroskedasticity-consistent Wald. In particular, we compute Davies' (1977, 1987) supremum test, which is the one we use in the main text, but we also compute Andrews and Ploberger's (1994) average and exponential average tests for each of the covariance matrices. These authors suggest that superior local power can be constructed by computing an average or an exponential average of the Wald test statistic (aveWald and expWald tests, respectively) over the parameter space admissible for k, the knot value.[25] We also include Wald and LM tests for the case where k is known and set to its true value. As a reminder, the moneyness of the option that best approximates hedge fund returns is not known a priori, so we include them only for comparison purposes. In particular, we find that the asymptotic approximation in Hansen (1996), which we use in our main text, delivers good size properties for the heteroscedasticity-consistent LM test: the proportion of rejections is around 5%

Table V. Finite sample size and power of asymptotic 5% size tests

Panel (a) $T = 60$

	Size	Power	
	$\delta = 0$	$\delta = \widehat{\delta}$	$\delta = 2\widehat{\delta}$
LM with k known	0.041	0.300	0.656
supLM	0.034	0.210	0.553
aveLM	0.049	0.239	0.599
expLM	0.055	0.255	0.598
Wald with k known	0.116	0.568	0.918
supWald	0.146	0.538	0.896
aveWald	0.129	0.490	0.869
expWald	0.106	0.430	0.825

Panel (b) $T = 100$

	Size	Power	
	$\delta = 0$	$\delta = \widehat{\delta}$	$\delta = 2\widehat{\delta}$
LM with k known	0.048	0.520	0.901
supLM	0.041	0.398	0.847
aveLM	0.048	0.419	0.856
expLM	0.055	0.416	0.849
Wald with k known	0.095	0.703	0.988
supWald	0.127	0.644	0.980
aveWald	0.100	0.613	0.974
expWald	0.086	0.560	0.950

Note: Simulations are based on the estimated parameters for the equally weighted index for category 9 (managed futures). All tests are robust to heteroskedastic errors.

[25] The asymptotic distribution of such a statistic can be computed by replacing step number (iv) in the simulation of the p-values with the corresponding (exponential) average of the random sample $\{T_n^1, \ldots, T_n^J\}$ of observations of the statistic.

DOI: 10.1002/jae

for a 5% size test, regardless of the sample size and choice of Monte Carlo design. However, the heteroscedasticity-robust Wald test tends to over-reject for our sample sizes: approximately 12% of rejections for the supWald test when the sample size is $T = 100$ and 15% when the sample size is $T = 60$. These results are consistent with the simulation study in Hansen (1996). Still, this size distortion is similar to the one we find for the Wald test when the position of the 'kink' is known and set to the true value: 10% when $T = 100$ and 12% when $T = 60$, and also similar to those found for the aveWald and expWald tests.

To assess finite-sample power, we set $\delta = \widehat{\delta}$ and $\delta = 2\widehat{\delta}$ while maintaining fixed the remaining parameters. As expected, the number of rejections (power) increases with δ and the sample size. In particular, we find that, for a sample size of $T = 100$, the proportion of rejections of supWald tests is close to 65% when $\delta = \widehat{\delta}$ and close to 98% when $\delta = 2\widehat{\delta}$. It is also worth mentioning that the finite-sample power of the Andrews and Ploberger (1994) average and exponential average tests tends to be smaller than their supremum test counterpart. Again, our results indicate that the loss of power with respect to the benchmark of a known knot set to its true value is small. For example, we find that the proportion of rejections when $\delta = 2\widehat{\delta}$ for the Wald test in which the knot is known and set to the true value is 98.8%, while the proportion of rejections for the supWald test is 98%.

5. CONCLUSION

We have shown that an approach to optimally searching for nonlinearities unveils strategies that look like put selling, straddles, or inverted straddles. However, given the limited information available on hedge fund returns, the statistical evidence is not as overwhelming as previous studies have concluded. Even if nonlinear strategies are employed, few categories provide a significantly positive value to investors, especially after accounting for the backfilling and liquidation biases. Quality funds can still be found in each category and our methodology helps identify them. Also, even if a fund delivers a null performance, it may still be of interest to an investor because it provides a nonlinear risk exposure that is otherwise not available from traded securities.

Our findings suggest that prudence should prevail when investing in hedge funds. Pension funds as well as retail investors have increased exposure to funds that engage in active and often difficult-to-decipher strategies. We hope that the tools developed in this paper will help investors recognize funds that offer the risk and return combination ultimately sought in these strategies. A challenging issue remains to see whether this characterization of nonlinear risk exposures is able to replicate hedge fund returns out-of-sample.

ACKNOWLEDGEMENTS

This paper started while the first author was a post-doctoral fellow at CIREQ and CIRANO; he is thankful for their hospitality. The second author gratefully acknowledges financial support from the Fonds Québécois de la recherche sur la société et la culture (FQRSC), the Social Sciences and Humanities Research Council of Canada (SSHRC), the Network of Centres of Excellence (MITACS), Hydro-Québec, and the Bank of Canada. Both authors also thank two anonymous referees, Greg Bauer, Georges Hübner, Michael King, Andrew Patton, Enrique Sentana, Marno Verbeek, Jun Yang, and seminar participants at the Bank of Canada, CEMFI, the 2005 CIREQ Conference on Time Series Models, and the Second Annual Empirical Asset Pricing Retreat at the University of Amsterdam for their comments. We also thank Vikas Agarwal and Narayan Y. Naik

DOI: 10.1002/jae

for kindly providing us with an updated time series of the returns on their option-based factors. The views expressed in this paper are those of the authors and do not necessarily reflect those of the bank of Canada.

REFERENCES

Ackermann C, McEnally R, Ravenscraft D. 1999. The performance of hedge funds: risk, return, and incentives. *Journal of Finance* **54**: 833–874.

Agarwal V, Naik NY. 2004. Risks and portfolio decisions involving hedge funds. *Review of Financial Studies* **17**: 63–98.

Agarwal V, Bakshi G, Huij J. 2007. *Higher-moment equity risk and the cross-section of hedge fund returns*. Mimeo, Georgia State University.

Amin GS, Kat HM. 2003. Hedge fund performance 1990–2000: do the 'money machines' really add value? *Journal of Financial and Quantitative Analysis* **38**: 251–274.

Andrews DWK, Ploberger W. 1994. Optimal Tests when a Nuisance Parameter is Present Only under the Alternative. *Econometrica* **62**: 1383–1414.

Chan N, Getmansky M, Hass SM, Lo A. 2007. Systemic risk and hedge funds. In *The Risks of Financial Institutions and the Financial Sector*, Carey M, Stulz R (eds). University of Chicago Press: Chicago, IL; 235–338.

Davies RB. 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64**: 247–254.

Davies RB. 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74**: 33–43.

Fung W, Hsieh DA. 2001. The risk in hedge fund strategies: theory and evidence from trend followers. *Review of Financial Studies* **14**: 313–341.

Glosten LR, Jagannathan R. 1994. A contingent claim approach to performance evaluation. *Journal of Empirical Finance* **1**: 133–160.

Hansen B. 1991. *Inference when a nuisance parameter is not identified under the null hypothesis*. Rochester Center for Economic Research Working Paper No. 296.

Hansen B. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* **64**: 413–430.

Hansen B. 1999. Testing for linearity. *Journal of Economic Surveys* **13**: 551–575.

Hansen LP, Richard SF. 1987. The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica* **55**: 587–614.

Hasanhodzic J, Lo AW. 2007. Can hedge-fund returns be replicated? The linear case. *Journal of Investment Management* **5**: 5–45.

Henriksson RD, Merton RC. 1981. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business* **54**: 513–533.

Hommel G, Hoffman T. 1988. Controlled uncertainty. In *Multiple Hypothesis Testing*, Bauer P, Hommel G, Sonnermann E (eds). Springer: Berlin; 154–161.

Lhabitant FS. 2004. *Hedge Funds: Myths and Limits*. Wiley: Chichester.

Malkiel B, Saha A. 2005. Hedge funds: risk and return. *Financial Analysts Journal* **61**: 80–88.

Mitchell M, Pulvino T. 2001. Characteristics of risk and return in risk arbitrage. *Journal of Finance* **56**: 2135–2175.

Patton AJ. 2007. Are 'market neutral' hedge funds really market neutral? *Review of Financial Studies* **22**: 2295–2330.

Posthuma H, van der Sluis PJ. 2003. A reality check on hedge fund returns. Mimeo, Free University Amsterdam.

Romano JP, Wolf M. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* **73**: 1237–1282.

Romano JP, Shaikh AM, Wolf M. 2008. Formalized data snooping based on generalized error rates. *Econometric Theory* **24**: 404–447.

Stulz RM. 2007. Hedge funds: past, present, and future. *Journal of Economic Perspectives* **21**: 175–194.

White HL. 2000. A reality check for data snooping. *Econometrica* **68**: 1097–1126.