Dynamic Hedge Fund Style Analysis with Errors in Variables

Laurent BODSON[1], Alain COËN[2] and Georges HÜBNER[3]

November 1, 2007

[1]Ph.D. candidate in finance and Research fellow FNRS at HEC-Management School, University of Liège (Belgium). Address: Rue Louvrex, 14 (N1) B-4000 Liège, Belgium | Phone: +32 4 2327431 | Fax: +32 4 2327240 | E-mail: Laurent.Bodson@ulg.ac.be

[2]Professor of finance at the Graduate School of Business (ESG), University of Quebec in Montreal (UQÀM) (Canada). Address: 315, rue Sainte-Catherine Est, H2X 3X2 (Canada) | Phone: (514) 987-3000 | Fax: (514) 987-0422 | E-mail: coen.alain@uqam.ca

[3]Deloitte professor of financial management at HEC-Management School, University of Liège, Associate professor of finance at the University of Maastricht, Academic expert at the Luxembourg School of Finance – University of Luxembourg, and Affiliate professor of finance at EDHEC. Address: Rue Louvrex, 14 (N1) B-4000 Liège (Belgium) | Phone: +32 4 2327428 | Fax: +32 4 2327240 | E-mail: G.Hubner@ulg.ac.be

Abstract

This paper revisits the traditional return-based style analysis (RBSA) in presence of time-varying exposures and errors in variables. We apply a selection algorithm using the Kalman filter to identify the more appropriate benchmarks and we compute their corresponding higher moment estimators (HME), i.e. the measurement error series introducing the (cross) moments of order three and four. Then, we retain the most significant HME and we add them to the selected benchmarks. Therefore, we obtain the most relevant benchmarks with none, some or all their HME as benchmarks explaining the analysed fund return. We finally run the Kalman filter on the principal components of this set of selected benchmarks to avoid multicollinearity problems. Analysing EDHEC alternative indexes styles, we show that this technique improves the factor loadings and permits to identify more precisely the return sources of the considered fund.

JEL Classification Codes: G11, G12 and C13

Keywords: Style Analysis; Kalman Filter; Errors in Variables; Higher Moment Estimators

1 Introduction

The investment style of a fund reveals its various sensitivities (or exposures) to a set of risk factors. This information is also often used for performance measurement. Return-based style analysis (RBSA) introduced by Sharpe [1988 and 1992] provides an estimate of the fund's historical exposures to style benchmarks. It consists of a constrained linear regression (with constant parameters) of the fund returns on relevant style index returns. The objective underlying style analysis is to set up a portfolio, composed of different tradable benchmarks, that shares the same exposures to systematic risk as the evaluated fund (see Lucas and Riepe [1996] among others).

Style-based analysis provides a convenient way to decompose the returns of managed portfolios into identifiable benchmarks and reproducible strategies. For this to be effective, it is very important to achieve a very high explanatory power with the return generating model for performance evaluation and risk-return analysis. First, the reliability of statistical inference on the portfolio alpha is contingent on the quality of the model. Second, a poor model specification leaves many sources of risk unexplained, making it difficult to reliably assess the risk and return characteristics of the funds.

Mutual funds and hedge funds do not present the same style characteristics. The transparency of mutual funds allows managers to apply them a holding-based or characteristic-based style analysis almost indifferently (see Brown and Goetzmann [2003] among others). Unfortunately, hedge fund operations are essentially opaque and qualitative assessments of investment styles are likely to be biased, and we have to turn to quantitative techniques (see Lhabitant [2004], pp.213-228). The explanatory power of RBSA applied to hedge funds appears to be limited. Indeed, the application of style analysis to hedge fund returns has met two types of hindrances that have not made it perfectly effective to date. On the one hand, managers of hedge funds are usually freer to modify their strategies than the ones of mutual funds. This results in time-varying risk sensitivities that cannot be captured by constant risk exposure coefficients. On the other hand, a large body of the extant literature contends that return-based indexes fail to account for option-based components in hedge fund returns, i.e. non-linear payoff patterns. There is indeed ample evidence (see Glosten and Jagannathan [1994], Fung and Hsieh [1997 and 2001], and Agarwal and Naik [2000]) that large gains in regression R-squared can be obtained by additional factors that use optional strategies. So far, these optional strategies have not yet been transformed into index-based return factors used in the pure RBSA, even though some proxies have been proposed in the literature. Moreover, the use of option-related benchmarks induces measurement error that may alter the regression specification and affect the accuracy and consistency of the estimated style exposures (see notably Samuelson [1970])

In this paper, we provide a framework aiming at jointly addressing the two issues of style analysis applied to hedge funds. In short, our framework complements the traditional RBSA with option-related benchmarks using a Kalman filter (selection) procedure with a correction for errors in variables (EIV).

Kalman filters[1] represent a rather natural econometric technique to meet the first difficulty listed above. They deal with style changes in a way that does not depend on arbitrarily chosen window sizes (Lhabitant [2004], pp.227-228). Indeed, the Kalman filter is suited to take into consideration the multiple investment style variations of actively managed funds (see Swinkels and Van Der Sluis [2006] and Corielli and Meucci [2004]), but its use has been rather limited in the literature due to the limited size of hedge fund databases[2].

To deal with EIV brought by the inflation of index benchmarks, we introduce additional instrumental regressors[3] that address the presence of measurement errors. These variables are computed with the (cross) sample moments of order three and four of the benchmarks initially selected. For this purpose, we use the technique presented by Dagenais and Dagenais [1997] to correct for errors in variables using instrument variables accounting for the higher moments of the selected benchmarks returns. Indeed, estimation applications incorporating higher moment estimators (HME) are particularly well suited to hedge fund return series (see Coën and Hübner [2006]), mainly when the implementation aims at characterizing dynamic variations of the exposures.

We also apply a factor analysis to the selected style benchmarks and HME in order to avoid multicollinearity problems and to identify the underlying risk factors. We finally run the Kalman filter on this ultimate factor series.

For these applications, we are confronted with a well-known, although seldom recognized paradox: inference based on the Kalman filter and factor analysis techniques is based on the normality assumption for the return series, while we explicitly acknowledge that we seek to capture non-linear

[1] The Kalman filter is an efficient recursive filter that estimates the state of a dynamic system from a series of noisy measurements.

[2] Capocci and Hübner [2004] have shown that data prior to 1994 is quite unreliable due to large survivorship and backfill biases, and should not be used for statistical inference.

[3] We only take the most significant ones.

behaviors in the return generating process. Therefore, our empirical application does not focus on the statistical properties of the estimators, but rather on their empirical predictive capacity. The acid test of the quality of our approach is thus related to the quality of the resulting predictions of returns.

This procedure appears to provide new insights on the returns of hedge fund strategies. As a result, we can evaluate the risk related to the analyzed fund but also its expected return. Knowing the expected returns of each selected benchmark and using the Kalman filter assumption that the loadings are locally stationary, i.e. we can use the current style exposures as predictors of the one period ahead style exposures, we have all the required information to compute the fund expected return. Out-of-sample tests on hedge fund indexes indicate a significant improvement over static (OLS) and dynamic (Kalman) versions of RBSA.

The paper proceeds as follows. In Section 2, we explain the methodology to identify the most relevant benchmarks of the RBSA with a selection algorithm using the Kalman filter. We introduce in Section 3 the Dagenais and Dagenais HME corresponding to the (cross) skewness and kurtosis of the selected benchmarks. Section 4 describes the Kalman filter algorithm applied to select the best HME combination added to the selected benchmarks. In Section 5, we compute the factors of the selected benchmarks and HME to eliminate potential multicollinearity between the explanatory variables. We run an ultimate Kalman filter on these factors and we transform the resulting estimates (valid for the factors) into exposures estimates for the selected benchmarks and HME. We finally compute the analysed fund expected return from the expected returns of each selected benchmark using the results of this composite model. Section 6 provides a backtesting study through an illustration. The last Section concludes.

2 Benchmark Selection

Selecting a wrong benchmark is one of the most frequent errors in portfolio management (see for instance Buetow et al. [2000]). Benchmarks are reference portfolios that simultaneously describe the diversity of risks taken by the fund during the analyzed period and define the fund's underlying management strategy. To avoid selection problems, we introduce an iterative procedure using the Kalman Filter and identifying the most relevant benchmarks using a principle related to stepwise regression.

Through style analysis, we want to dynamically replicate the fund return by a portfolio composed of selected benchmarks. On the one hand, we impose that the sum of the benchmark weights in the replicating portfolio is equal to one in order to give them the interpretation of portfolio weights. Furthermore, the loadings must be positive or null because we do not allow short selling of the selected benchmarks (Ter Horst, Nijman and De Roon [2004]). On the other hand, Fung and Hsieh [1997] and Brown and Goetzmann [2003], among others, report that the sensitivities of hedge funds to style exposures are time-varying. To account for this property, we compute the following constrained Kalman filter (with (1) being the measurement equation, (2) being the transition equation and (3) the two constraints) to deduce the exposure values at each time t of the historical series, i.e. for $t = 0, \ldots, T$ (see Doran [1992]):

$$
\begin{aligned}
R_t &= w_{0,t} + \sum_{i=1}^{k} w_{i,t}.R_t^i + \varepsilon_t, && (1)\\
w_{j,t+1} &= w_{j,t} + \zeta_{j,t+1}, \text{ for } j = 0, ..., k && (2)\\
\sum_{i=1}^{k} w_{i,t} &= 1 \text{ and } w_{i,t} \geq 0, \text{ for } i = 1, ..., k && (3)
\end{aligned}
$$

where R_t represents the fund's historical return at time t, $w_{0,t}$ is the time-varying intercept of the RBSA at time t, k is the number of selected benchmarks (asset class factors), R_t^i represents the historical return of the benchmark i at time t, ε_t reflects idiosyncratic noise at time t and $w_{i,t}$ represents the weight (or exposure) of the benchmark i at time t. In addition, the error terms are distributed according to (4).

$$
\varepsilon_t \backsim NID(0, \sigma_\epsilon^2) \text{ and } \zeta_{j,t} \backsim NID(0, \sigma_{j,\zeta}^2), \; j = 0, ..., k. \qquad (4)
$$

Building on this traditional Kalman filter, we propose a selection procedure that follows similar principles to the linear stepwise regression analysis. The procedure consists of starting with a very large number of (potentially adequate) benchmarks ($> k$) and running the Kalman filter presented above using the benchmarks taken separately, i.e. one by one. Initially, we find the first benchmark that minimizes the mean-squared error (MSE) of the Kalman filter[4]. The MSE of the Kalman filter is simply the mean-squared

[4] The R-squared is not a valid fitting measure in this case because we do not use an ordinary least squares approximation but a constrained Kalman filter.

difference between the return estimate from the Kalman filter and the true return at each time t (therefore $\frac{1}{T+1} \cdot \sum_{t=0}^{T} \varepsilon_t^2$). This is a measure of the fitting power of the Kalman filter with the selected benchmarks. We repeat the procedure to select a second benchmark (from the remaining benchmarks) minimizing the MSE of the new Kalman filter integrating the first benchmark chosen. We proceed with this incremental procedure until we have the desired number k of selected benchmarks. In analogy with the heuristic stopping rule with many PCA approaches, we stop the procedure, and therefore fix the parameter k, when the improvement of the optimal Kalman filter MSE is less than 5%. Eventually, these k benchmarks correspond to the style exposures of the analyzed fund and explain the greatest part of its return behavior over time. This results in a k-factor model which is used to determine the fund asset mix. Note that the benchmarks selected for a specified fund can change from one period to another.

3 Errors in Variables

Once the benchmarks are selected, we want to create the corresponding HME (considered as additional benchmarks). We use the method developed by Dagenais and Dagenais [1997] and applied by Coën and Hübner [2006] thereby creating new regressors accounting for the estimated measurement errors using the higher moments, i.e. the (cross) skewness and the (cross) kurtosis. To compute the measurement error series corresponding to each selected benchmark, we use the following formulas (6 to 8) and we run k (artificial) ordinary least squares regressions (5):

$$SB = (i, z_1, z_2).\hat{\Lambda} + \hat{W} \tag{5}$$

with

$$\begin{aligned} f &= (I_n - \frac{ii'}{n}).SB, & (6) \\ z_1 &= f * f, & (7) \\ z_2 &= f * f * f - 3f.(E[\frac{f'f}{n}] * I_k) & (8) \end{aligned}$$

where n is the number of observations, i is a $(n \times 1)$ vector where the elements are all one, I_n and I_k are identity matrix of order n and k respectively, SB denotes a $(n \times k)$ matrix containing the k selected benchmarks, $*$ is the Hadamard element-by-element matrix multiplication operator, the matrix

f stands for the matrix SB calculated in mean deviation, $\hat{\Lambda}$ is a $(n \times k)$ matrix containing estimators and $\hat{W}$ is a $(n \times k)$ matrix standing for the estimated matrix of error terms. We run k artificial regressions to have k series of error terms (the k columns of $\hat{W}$). In addition, we know from the least squares approximation that the mean of the residual series is null, i.e. $E[\hat{W}] = 0$.

4 Higher Moment Estimator Selection

We have initially k selected benchmarks with their k corresponding measurement error series. We run the Kalman filter on the k selected benchmarks plus each possible combination of the error series to compute the MSE of each combination[5]. All the possible selections can be enumerated because it is simply a combination without repetition of i (for $i = 0, \ldots, k$) elements chosen from k elements (see combinatorial analysis). We run the following constrained Kalman filter (with (9) being the measurement equation, (10) being the transition equation and (11) the three constraints) for each possible combination and for $t = 0, \ldots, T$:

$$R_t = w_{0,t} + \sum_{i=1}^{k} \left(w_{i,t}.R_t^i + w_{k+i,t}.W_t^i \right) + \varepsilon_t, \tag{9}$$

$$w_{j,t+1} = w_{j,t} + \zeta_{j,t+1}, \text{ for } j = 0, ..., 2k \tag{10}$$

$$\sum_{i=1}^{k} w_{i,t} = 1 \ , \sum_{i=k+1}^{2k} w_{i,t} = 0 \text{ and } w_{i,t} \geq 0, \text{ for } i = 1, ..., k \tag{11}$$

where R_t represents the fund's historical return at time t, $w_{0,t}$ is the time-varying intercept of the RBSA at time t, k is the number of selected benchmarks, R_t^i represents the historical return of the benchmark i at time t, W_t^i corresponds to the historical measurement error series of the benchmark i at time t (computed in Section 3), ε_t reflects idiosyncratic noise at time t, $w_{i,t}$ represents the weight of the benchmark i at time t and $w_{k+i,t}$ represents the measurement error series of the benchmark i at time t. The value of $w_{k+i,t}$ is set to zero if the measurement error series related to the benchmark i is not selected.

Note that the constraints corresponding to the selected benchmarks are not modified. However, the unique constraint established for the measure-

[5]We can select no, some or all error series.

ment error series imposes that the sum of their weights at each time t is null.

We save the combination that minimizes the MSE of the Kalman filter. The final number of benchmarks k_f (i.e. the number of selected benchmarks k plus the number of selected measurement error series $k_f - k$) is comprised between k and $2k$ (included).

5 Predictive Capacity

Like for OLS estimation, our constrained dynamic style analysis produces estimators based on the variance-covariance matrix of the regressors.[6] Unfortunately, given the large number of candidate benchmarks and HME, many of them are likely to exhibit serious multicollinearity. Even though this issue does not affect the explanatory power of the model, it is likely to influence the stability of the regression coefficients as well as the quality of statistical inference based thereupon, such as with the procedure proposed by Lobosco and DiBartolomeo [1997]. Such issues may affect the quality of return predictions based on the model, and thus the generation of conditional expected returns based on the Kalman filter analysis.

In order to avoid these multicollinearity problems (see Agudo and Gimeno [2005] among others) for the next steps in the procedure, we express the selected benchmarks and HME data in terms of factors. This type of analysis allows us to detect the structure in the relationships between the explanatory variables. Note that we do not reduce the dimension of the data because we need to come back to the initial dataset for explanatory convenience. We firstly establish the variance-covariance matrix of the selected benchmarks and the selected HME. We compute then the eigenvectors of this variance-covariance matrix. The corresponding eigenvalues are not required because we do not want to reduce the data dimension. The factor matrix of the selected benchmarks and the associated HME, noted Φ, is computed with the following formula:

$$\Phi = (eigenvectors'.D')' \quad (12)$$

where $eigenvectors$ is a $(k_f \times k_f)$ matrix columns represent the eigenvectors, D is a $(n \times k_f)$ matrix containing the k selected benchmarks returns series

[6] We apply this procedure even though we may be lead to empirically reject the normality hypothesis of the return series. This is the reason why, in this paper, we do not focus on the statistical significance of the coefficients, but rather on the predictive capacity of the model.

and the $k_f - k$ selected HME series sorted by column and Φ is a $(n \times k_f)$ matrix containing the k_f orthogonal factors which are linear combinations of the selected benchmarks returns and the selected HME (see Corielli and Meucci [2004]). We get back to the initial data using the inverse transformation formula because we keep all the factors or, in other words, all the eigenvectors.

Then, we run an ultimate Kalman filter on these factors to obtain the exposures of the fund to the factors at each time t ($\gamma_{i,t}$ for $i = 1, \ldots, k_f$). We impose three constraints on these exposures in order to respect the non negativity and the unitary sum assumptions for the selected benchmarks weights ($w_{i,t}$ for $i = 1, \ldots, k$) and the null sum for the selected measurement error series (or HME) weights ($w_{i,t}$ for $i = k+1, \ldots, k_f$) at each time t. Thereby, we run the following Kalman filter (13 and 14) for $t = 0, \ldots, T$ with the following three constraints (15 and 16):

$$R_t = \gamma_{0,t} + \sum_{i=1}^{k_f} \gamma_{i,t}.\Phi_t^i + \varepsilon_t, \tag{13}$$

$$\gamma_{j,t+1} = \gamma_{j,t} + \zeta_{j,t+1}, \text{ for } j = 0, ..., k_f \tag{14}$$

$$(eigenvectors'.V_B)'.\Gamma_t = 1 \text{ and } (eigenvectors'.V_{HME})'.\Gamma_t = 0 \tag{15}$$

$$eigenvectors.\Gamma_t \geq K \tag{16}$$

where Φ_t^i stands for the column i of Φ at time t, V_B represents a $(k_f \times 1)$ column vector of k ones then $k_f - k$ zeros, V_{HME} represents a $(k_f \times 1)$ column vector of k zeros followed by $k_f - k$ ones, Γ_t is the column vector of the $\gamma_{i,t}$ for $i = 1, \ldots, k_f$ and K is a column vector of order k_f with k zeros followed by $k_f - k$ minus infinity. As previously, we do not constraint $\gamma_{0,t}$, the time-varying intercept of the RBSA at time t.

In order to convert the $\gamma_{i,t}$ to real exposures of the fund to the selected benchmarks and HME, i.e. the $w_{i,t}$ (for $i = 1, \ldots, k_f$), we need to use the following transformation equation:

$$weights = (eigenvectors.\Gamma)' \tag{17}$$

where *weights* is a $(n \times k_f)$ matrix containing the weight (or exposure) estimates of each selected benchmark and HME at each time t, *eigenvectors* is a $(k \times k_f)$ matrix whose columns represent an eigenvectors (identical to the matrix deduced supra) and Γ stands for the $(k_f \times n)$ matrix with the exposures of the fund to the factors at each time t ($\gamma_{i,t}$ for $i = 1, \ldots, k_f$).

Finally, we can compute the fund return in one period using the estimated current exposures ($w_{i,T}$ and $w_{k+i,T}$) if we assume that they are identical to those in one period ($w_{i,T+1}$ and $w_{k+i,T+1}$). Indeed, the transition equation of the Kalman filter justifies this assumption because the process noise is assumed to be drawn from a zero mean multivariate normal distribution. In other words, the Kalman assumes that $E_t(w_{i,t+1}) = w_{i,t}$. Concerning the independent term $\gamma_{0,T+1}$, we suggest to compute an average over all the sample ($E[\gamma_{0,t}]$ for $t = 0, \ldots, T$) or an average over the "last" period's because this parameter exhibits larger movements than the exposures. Knowing the expected return of each selected benchmark, we have all the information to compute the fund expected return equation:

$$E[R_{T+1}] = \gamma_{0,T+1} + \sum_{i=1}^{k} \left(w_{i,T+1}.E[R^i_{T+1}] + w_{k+i,T+1}.E[W^i_{T+1}] \right) \tag{18}$$

where $w_{i,T+1}$ and $w_{k+i,T+1}$ represent the future exposures of the fund to the benchmark i or its measurement error respectively. As previously mentioned, the value of $w_{k+i,T+1}$ is set to zero if the measurement error series related to the benchmark i is not selected. Furthermore, we know that $E[W^i_{T+1}]$ is equal to zero by construction (see Section 3), we have therefore the following simplified expression of the fund expected return:

$$E[R_{T+1}] = \gamma_{0,T+1} + \sum_{i=1}^{k} w_{i,T+1}.E[R^i_{T+1}]. \tag{19}$$

This predictor can be compared against the ones produced by a constrained static style analysis as in Sharpe [1992] and a dynamic Kalman filter analysis similar to Swinkels and Van Der Sluis [2006] with stepwise benchmark selection.

6 Empirical Application

To illustrate the usefulness of our approach, we consider the 14 EDHEC alternative indexes for the 2000-2006 period, a total of 78 months. The Table 1 exhibits a summary statistics of the indexes monthly returns[7]. The columns contain information about the mean, the standard deviation, the

[7] These data are available on www.edhec-risk.com.

skewness (standardized), the excess kurtosis (standardized), the minimum, and the maximum of the monthly returns[8].

Table 1: EDHEC Alternative Indexes Monthly Returns - Descriptive statistics

Hedge Fund Strategy	Mean	Std. Dev.	Skewness	X-Kurtosis	Min	Max
Convertible Arbitrage	0.0059	0.0107	-0.5134	1.2030	-0.0316	0.0344
CTA Global	0.0066	0.0273	0.0010	-0.4275	-0.0543	0.0682
Distressed Securities	0.0106	0.0116	-0.0099	-0.1787	-0.0209	0.0360
Emerging Markets	0.0112	0.0239	-0.5547	-0.3280	-0.0462	0.0586
Equity Market Neutral	0.0055	0.0045	0.2466	1.4167	-0.0082	0.0210
Event Driven	0.0080	0.0125	-0.7462	1.1287	-0.0300	0.0341
Fixed Income Arbitrage	0.0056	0.0047	-0.0605	1.7290	-0.0092	0.0207
Global Macro	0.0070	0.0124	0.5557	0.4386	-0.0178	0.0472
Long/Short Equity	0.0057	0.0172	-0.3952	-0.4800	-0.0389	0.0381
Merger Arbitrage	0.0050	0.0085	-1.0734	2.3275	-0.0267	0.0272
Relative Value	0.0062	0.0089	-0.3346	1.3511	-0.0221	0.0333
Short Selling	0.0048	0.0476	0.5073	0.9912	-0.1135	0.1657
Funds of Funds	0.0051	0.0105	-0.1722	-0.4179	-0.0205	0.0286
Multi Strategy	0.0076	0.0075	0.1129	0.4114	-0.0137	0.0304

As acknowledged by a growing literature, the two first moments are insufficient to provide a good description of risk. The descriptive statistics (reported in Table 1) and the normality tests that we performed on each variable[9] support this view. Consequently, we can reject the hypothesis of returns normality. This suggests that higher moments of the regressors are highly likely to influence the EDHEC indexes performance measurement.

We consider a large set of 118 potential style benchmarks selected among the literature (reported in the Appendix). We take the major equity, bond, commodity, real estate, exchange and interest indices over all the continents (see Capocci, Corhay and Hübner [2005] and Fung and Hsieh [2002] among others). We also incorporate the Fama and French factors [1993] and the additional momentum factor proposed by Carhart [1997]. We take into consideration the potential impact of hedge fund stale prices (and return smoothing) by taking lagged values (one month and two months) of several main benchmarks (see Okunev and White [2006] and Getmansky, Lo and

[8] We assume to have a population and not a sample of the returns, i.e. we do not use a correction term to compute the standard deviation, the skewness and the excess kurtosis.

[9] The results from the Jarque-Bera, Lilliefors, Anderson-Darling, Cramer-von Mises, Watson and Anderson-Darling tests are available on demand.

Makarov [2004]). We compute several spreads on different indices according to Fung and Hsieh [2002].

We select some particular benchmarks as the coskewness and the cokurtosis with basic benchmarks (see Lhabitant [2004] p.199 and Kraus and Litzenberger [1976])[10], the US implied volatility index VIX (see Hübner and Papageorgiou [2006]) and some swap instruments (see Okunev and White [2006] and Agarwal and Naik [2000]).

We include option-based factors computing at-the-money, out-of-the-money and in-the-money put and call options on several major indices (see Okunev and White [2006], Hübner and Papageorgiou [2006], Agarwal and Naik [2000], Henriksson and Merton [1981] and Glosten and Jagannathan [1994]). The moneyness is fixed at 5% and the maturity is equal to one month according to the results empirically derived by Diez de los Rios and Garcia [2005]. Finally, we use the Fung-Hsieh lookback straddles (see Fung and Hsieh [2001]) and we create other synthetic look-back straddles on different indices (see Fung and Hsieh [2002] and Goldman, Sosin and Gatto [1979]). Indeed, these (nonlinear) instruments explain a large part of the nonlinear patterns of the EDHEC indexes returns.

In order to validate the composite model, we compare the results accuracy obtained according to (a) the traditional method, i.e. a multiple linear regression (an ordinary least squares) with the two constraints of non negativity and unitary sum of the exposures, (b) a general Kalman filter run with the same constraints and (c) our composite Kalman-HME-factor model. We test a walk-forward using the usual technique of out-of-sample forecasting. We take a sliding window of 30 months and we repeat the process estimation moving forward the window until there are no more data to test. The results accuracy of the three models is judged based on the out-of-sample test results, i.e. the prediction error of the considered model.

The following three tables contain the results from the OLS (Table 2), the general Kalman filter (Table 3) and the composite (Kalman-HME-factor) model (Table 4) on all the potential benchmarks. The MSE here expresses the difference between the expected return estimate from the considered model and the true expected return for each sub-period of the sliding window. We expose for each model the root mean-squared error (RMSE) which is simply the squared root of the MSE, the minimum, the maximum and the standard deviation of the prediction errors.

[10] We could have included some other moment-related variables (for instance 'high skew minus low skew' portfolio returns) ensuring that these benchmarks are tradable.

Table 2: Results from the OLS

Hedge Fund Strategy	RMSE	Min	Max	Std. Dev.
Convertible Arbitrage	0.0231	-0.0454	0.0667	0.0222
CTA Global	0.0422	-0.0751	0.1311	0.0418
Distressed Securities	0.0251	-0.0482	0.0594	0.0242
Emerging Markets	0.0284	-0.0697	0.0812	0.0284
Equity Market Neutral	0.0293	-0.0648	0.0996	0.0286
Event Driven	0.0434	-0.0938	0.1280	0.0431
Fixed Income Arbitrage	0.0254	-0.0487	0.0686	0.0253
Global Macro	0.0359	-0.0884	0.0866	0.0358
Long/Short Equity	0.0321	-0.0705	0.1021	0.0321
Merger Arbitrage	0.0270	-0.0549	0.0851	0.0270
Relative Value	0.0278	-0.0626	0.1110	0.0277
Short Selling	0.0423	-0.0980	0.0978	0.0420
Funds of Funds	0.0282	-0.0609	0.0856	0.0279
Multi Strategy	0.0234	-0.0612	0.0536	0.0232

Table 3: Results from the General Kalman Filter

Hedge Fund Strategy	RMSE	Min	Max	Std. Dev.
Convertible Arbitrage	0.0188	-0.0137	0.0492	0.0113
CTA Global	0.0286	-0.0605	0.0622	0.0278
Distressed Securities	0.0136	-0.0231	0.0282	0.0112
Emerging Markets	0.0220	-0.0455	0.0503	0.0217
Equity Market Neutral	0.0051	-0.0116	0.0115	0.0051
Event Driven	0.0125	-0.0216	0.0278	0.0124
Fixed Income Arbitrage	0.0063	-0.0063	0.0135	0.0048
Global Macro	0.0140	-0.0286	0.0277	0.0138
Long/Short Equity	0.0181	-0.0397	0.0326	0.0172
Merger Arbitrage	0.0091	-0.0235	0.0170	0.0090
Relative Value	0.0093	-0.0165	0.0232	0.0087
Short Selling	0.0349	-0.0871	0.0783	0.0349
Funds of Funds	0.0110	-0.0269	0.0239	0.0110
Multi Strategy	0.0098	-0.0182	0.0269	0.0089

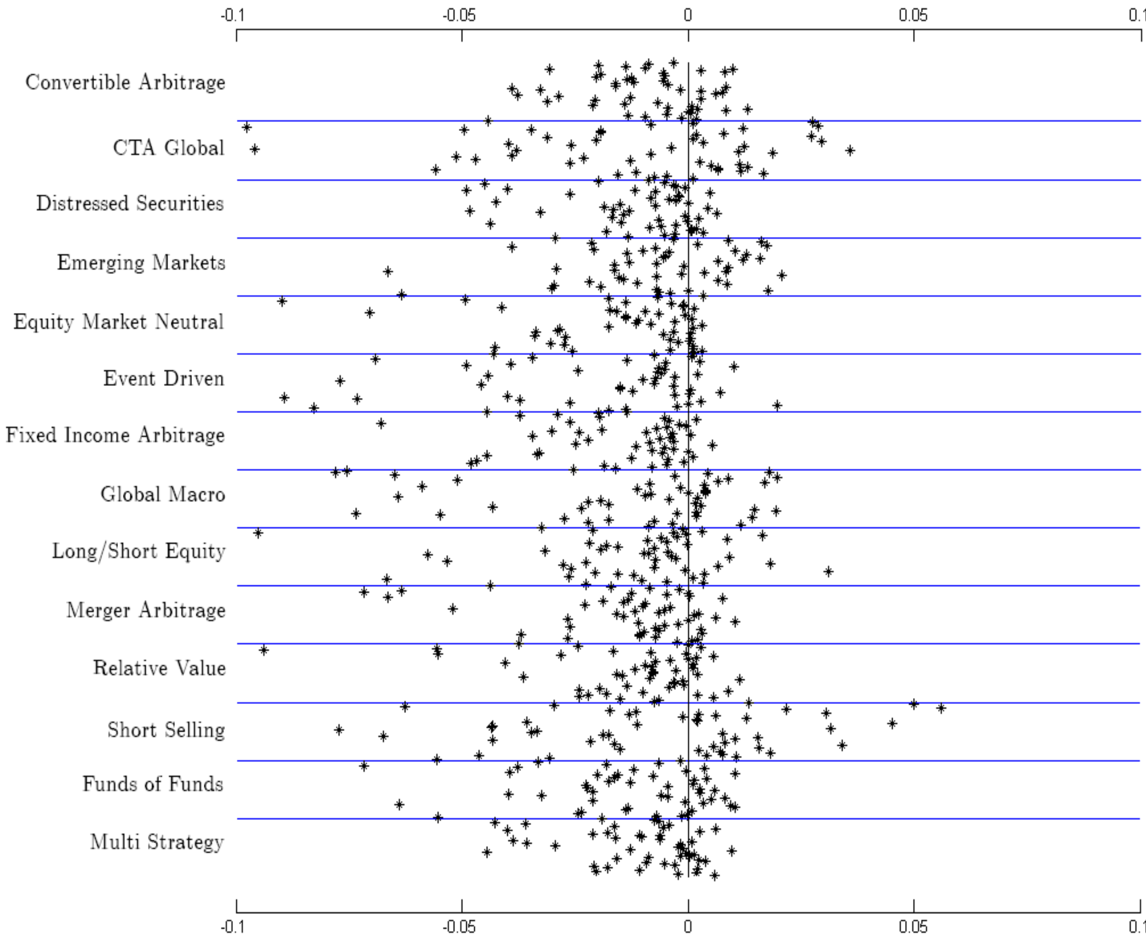
Table 4: Results from the Composite Model

Hedge Fund Strategy	RMSE	Min	Max	Std. Dev.
Convertible Arbitrage	0.0113	-0.0244	0.0300	0.0113
CTA Global	0.0267	-0.0558	0.0633	0.0267
Distressed Securities	0.0102	-0.0265	0.0136	0.0101
Emerging Markets	0.0205	-0.0426	0.0515	0.0205
Equity Market Neutral	0.0049	-0.0092	0.0134	0.0049
Event Driven	0.0232	-0.0829	0.1134	0.0231
Fixed Income Arbitrage	0.0048	-0.0126	0.0160	0.0048
Global Macro	0.0139	-0.0291	0.0259	0.0138
Long/Short Equity	0.0161	-0.0398	0.0324	0.0158
Merger Arbitrage	0.0079	-0.0259	0.0162	0.0077
Relative Value	0.0078	-0.0208	0.0170	0.0078
Short Selling	0.0334	-0.0837	0.0724	0.0323
Funds of Funds	0.0104	-0.0253	0.0179	0.0103
Multi Strategy	0.0084	-0.0222	0.0177	0.0083

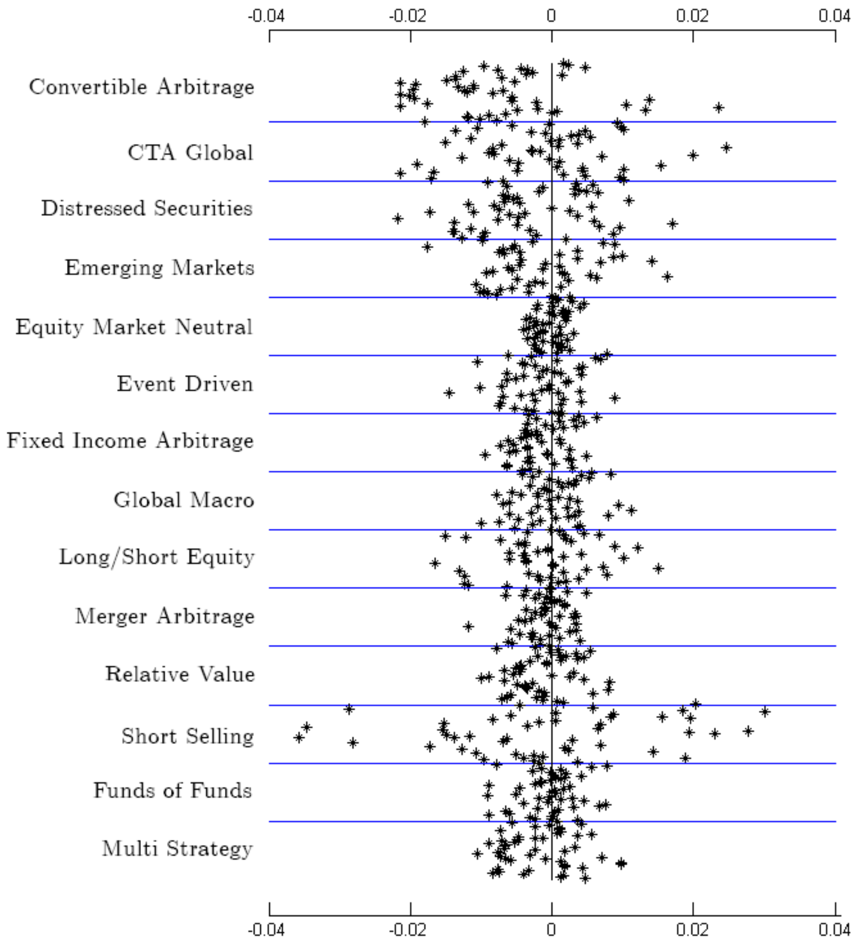
We observe in these results tables that the composite model accuracy exceeds on average the accuracy of the style analysis using the OLS or the general Kalman filter. Indeed, all the parameters assessing the predictive power of the three approaches present the composite model as globally superior. We note that the "interval error" (the difference between the maximum error and the minimum error) of the OLS is larger than the "interval error" of the other two specifications. Furthermore, the mean standard deviation of the OLS and the general Kalman filter is higher than the composite model mean standard deviation.

To investigate our remarks, we plot for each test (48 tests for each EDHEC alternative index, i.e. 672 predictions per method), the difference between the absolute prediction error of the composite model and the absolute prediction error of the OLS (Graph 1) or the general Kalman filter (Graph 2). All the points lying on the left (respectively on the right) of the vertical axis (representing a difference equal to zero between the predictive power of each model) mean that the composite model is more (respectively less) accurate in predicting the expected return then the OLS or the general Kalman filter.

Graph 1: The Composite Model Absolute Prediction Error minus the OLS Absolute Prediction Error (by Hedge Fund Strategy)

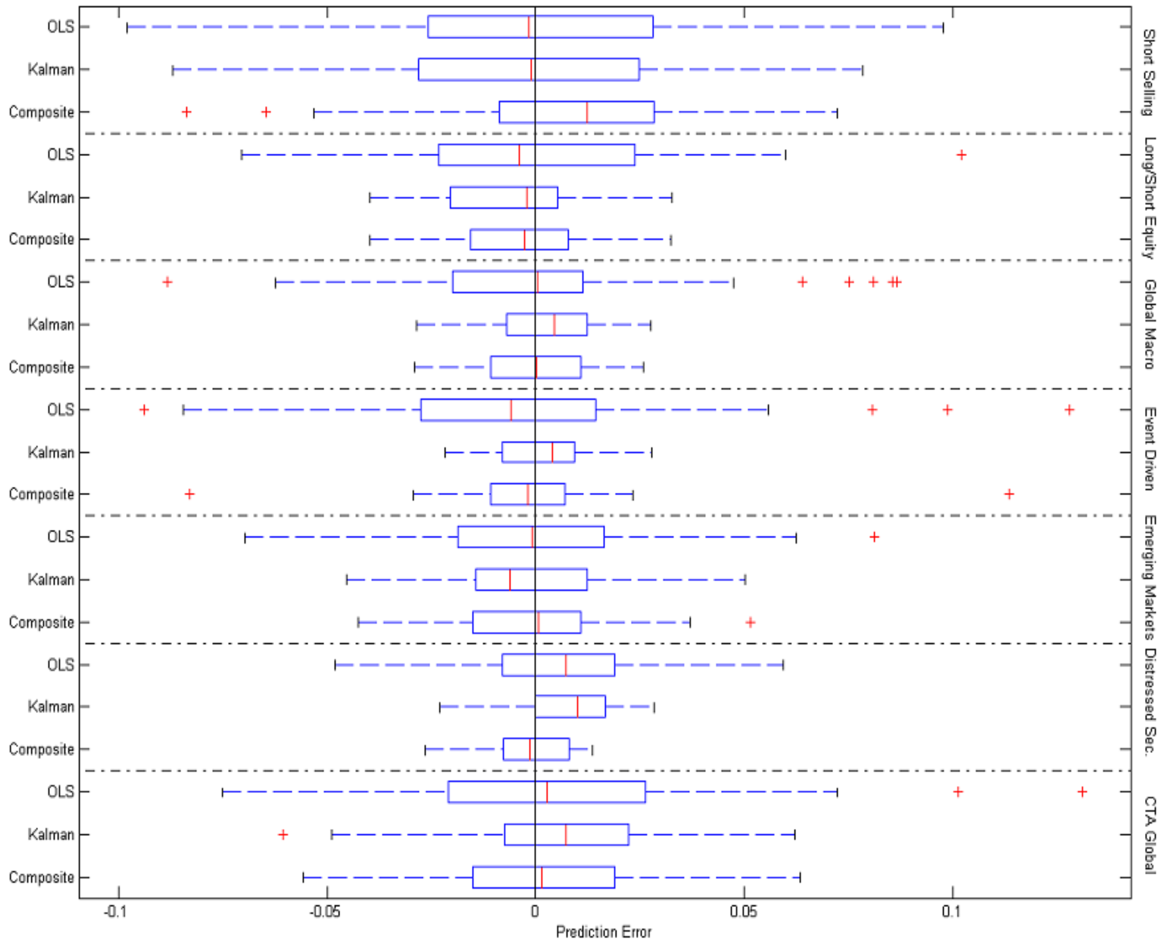


Graph 2: The Composite Model Absolute Prediction Error minus the General Kalman Filter Absolute Prediction Error (by Hedge Fund Strategy)

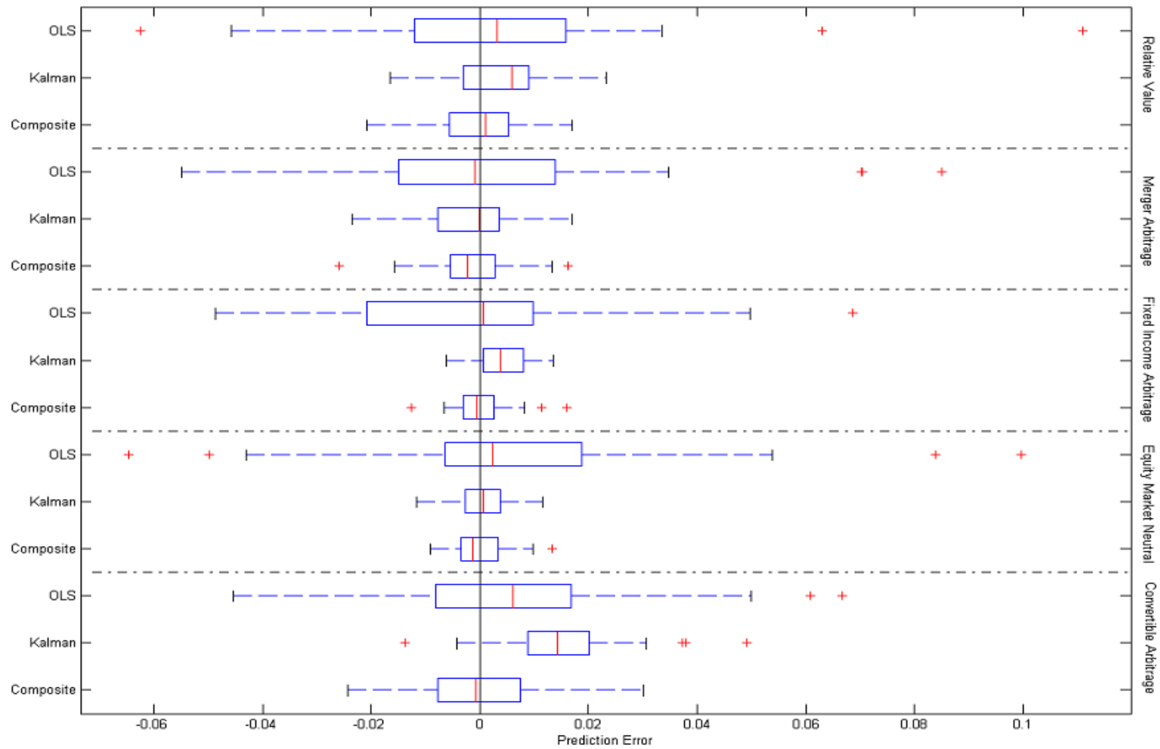


We observe from the graphs that the majority of the points (70.98% for Graph 1 and 59.97% for Graph 2) lie on the left of the vertical axis meaning that the composite model performs on average better in predicting the future returns of the analyzed EDHEC indexes. However, we easily note that some strategies show better results than others. To investigate this point, we plot, for each model prediction, the box-and-whisker diagram to indicate graphically the statistical distribution. We categorize the 14 EDHEC alternative indexes in three main groups: the directional (Graph 3), the non-directional (Graph 4) and the other strategies (Graph 5).

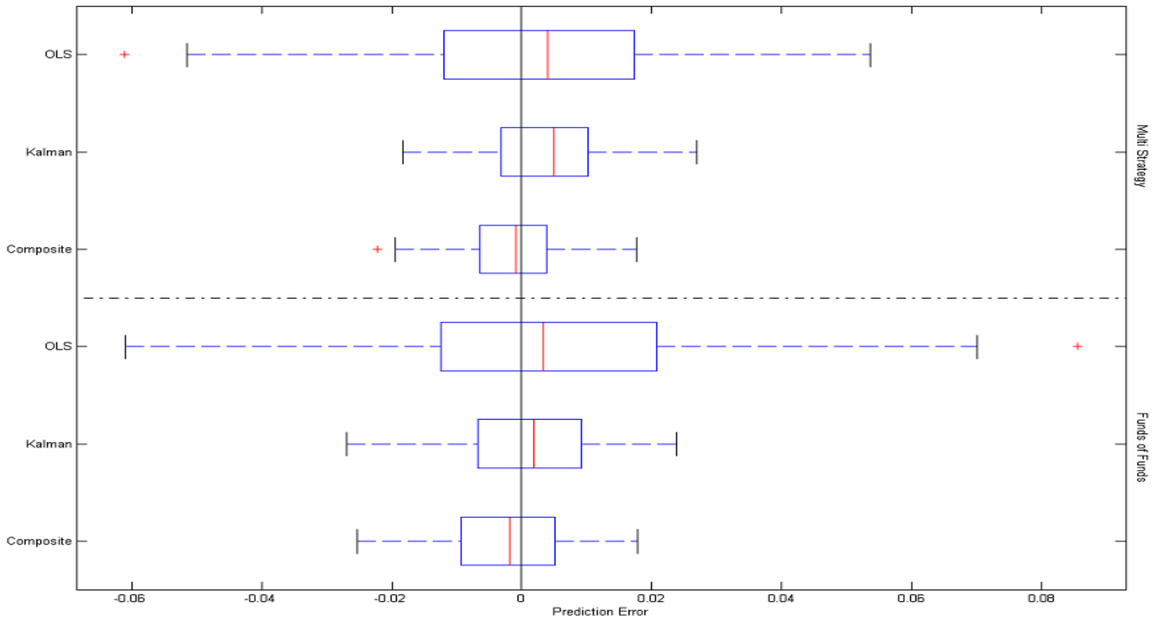
Graph 3: Prediction Error of The Directional Strategies (According to Each Model)



Graph 4: Prediction Error of The Non-Directional Strategies (According to Each Model)



Graph 5: Prediction Error of The Other Strategies (According to Each Model)



We note that the CTA Global, Emerging Markets, Long/Short Equity and Short Selling strategies have large interquartile range and whiskers. The mean prediction error is closer to zero with the composite index for all strategies except for the Short Selling strategy. On the one hand, the Emerging Markets index is the one that displays the largest variability of returns, and the CTA index is known to display the poorest fit with a linear style analysis. These outlying behaviors probably explain the relatively poor fit of the procedure. On the other hand, it sounds obvious that the two strategies integrating short positions do not meet the positivity constraint for the selected benchmarks weights. Indeed, if we impose that the weights are negative or null for Short Selling index and if we control the weights between -1 and 1 for the Long/Short Equity index, we would obtain better results.

Overall, these results emphasize the improvement in the exposures estimates brought by the composite model through its selection procedure and higher moment incorporation. The applications of this type of analysis can be useful for the determination of the (time-varying) expected returns of different strategies. Obviously, this type of analysis should be applied to individual funds for screening and allocation decisions. The possibility to disaggregate the analysis at the single fund or portfolio level is the major challenge to be met by our composite approach.

7 Conclusions

The main contribution of this article is to propose and illustrate a comprehensive technique to identify the time-varying exposures of a fund to different appropriate benchmarks and measurement errors. The Kalman filter utilization and the higher moment estimators proposed by Dagenais and Dagenais [1997] produce increases in the explanatory power of the RBSA (Sharpe [1988 and 1992]) for the sample test that we analyzed. Specifically, we compared the prediction error of the composite model to the prediction error of two ordinary methods using the EDHEC alternative indexes. Through two main selection algorithm (the selection of the benchmarks and the selection of the HME), we compute the expected return series of each EDHEC index.

The findings of this article suggest that the model developed improves the accuracy of the traditional prediction using the RBSA. The selection algorithm permits to avoid irrelevant benchmarks and the factor analysis obviates the potential multicollinearity between the selected benchmarks. Further research should focus on obtaining more accurate initial benchmarks. In addition, the algorithm that we propose brings other improvements for the measurement error series selection.

The application scope of our approach seems to be very large. On the one hand, the risk management can identify the risk exposures of a specified fund but also the effect of the benchmarks higher moments. On the other hand, the management can optimize its portfolio using the expected returns provided by the model.

We view our contribution as a first step towards dynamic style analysis in presence of errors in variables and faced with a huge number of potential benchmarks. We aim at improving the benchmark (or asset class) selection, the time-varying exposure estimates to the different risk sources (including the higher moments) and, ultimately, the assessment of the fund expected return.

Appendix: Benchmarks by Asset Class

Equity

RUSSELL 3000, S&P 500 COMPOSITE, NASDAQ 100, DJ STOXX EUR, MSCI EUROPE, NIKKEI 300, MSCI EM, MSCI WORLD EX US

High Yield

ML US HIGH YIELD MASTER II.

Bond

LEHMAN MUNICIPAL BOND, JPM EUROPE GOVT BOND, JPM US GOVT.BOND, JPM JAPAN GOVT.BOND, LEHMAN US AGGREGATE CORP BAA, LEHMAN US CREDIT BOND INDEX

Commodity

S&P GSCI Commodity, S&P GSCI Precious Metal, S&P GSCI Agriculturl

Mortgage

LEHMAN MBS HYBRID ARM

Real Estate

DJTM WESTERN EUROPE REAL ESTATE, DJTM WORLD REAL ESTATE, FTSE W US REAL ESTATE

Swap

SWAP Index

Lag 1 and 2

RUSSELL 3000, S&P 500 COMPOSITE, NASDAQ 100, SMB, HML, WML, DJ STOXX EUR, MSCI EUROPE, NIKKEI 300, ML ASIA/ PACIFIC CONV.BND, MSCI EM

Interest Rate

US TREASURY COMPOSITE >10 YR, US TREAS.BILL 3M, FED FUND EFFECTIVE, LEHMAN US TREASURY: 7-10 Y

Fama French

SMB and HML

Carhart

WML

Co-moment

Co-Skewness and Co-kurtosis

Spread

ML US HIGH YIELD MASTERII., LEHMAN US AGG. CORPORATE BAA, FHA, CONV MORTGAGE, SWAP RATE, ML GLOBAL 300 CONVERTIBLE, CITIGROUP.WORLDGOVT.BOND, Moody's Baa - US Treasury 10y

Option (Put/Call atm and otm)

S&P500, RUSSELL3000, Nasdaq100, MSCI World EX.US, VIX, STRADDLE HY-TRSY, STRADDLE CDY-TRSY, CBOE VIX

Fung-Hsieh factors

Bond lookback straddle, Currency Lookback Straddle, Commodity Lookback Straddle, Short Term Interest Rate Lookback Straddle, Stock Index Lookback Straddle

Currency

Trade Weighted Exchange Index: Broad

References

[1] Agarwal, V. and Naik, N.Y. (2000) Performance Evaluation of Hedge Funds with Option-Based and Buy-and-Hold Strategies. London Business School, *Working paper.*

[2] Agudo, L.F. and Gimeno, L.A.V. (2005) Effects of Multicollinearity on the Definition of Mutual Funds' Strategic Style: the Spanish Case. *Applied Economics Letters*, 12, 553-556.

[3] Brown, S.J. and Goetzmann, W.N. (2003) Hedge Funds With Style. *Journal of Portfolio Management*, 29, 101-112.

[4] Buetow, J., Gerald, W., Johnson, R.R. and Runkle, D.E. (2000) The Inconsistency of Return-Based Style Analysis. *Journal of Portfolio Management*, 26, 61-77.

[5] Capocci, D., Corhay A. and Hübner, G. (2005) Hedge Fund Performance in Bull and Bear Markets. *European Journal of Finance*, 5, 361-392.

[6] Carhart, M. (1997) On Persistence in Mutual Funds Performance. *Journal of Finance*, 52, 57-82.

[7] Coën, A. and Hübner, G. (2006) Risk and Performance Estimation in Hedge Funds: Evidence from Errors in Variables. HEC-University of Liège, *Working paper.*

[8] Corielli, F. and Meucci, A. (2004) Pitfalls in Linear Models for Style Analysis. *Statistical Methods and Applications*, 13, 103-127.

[9] Dagenais, M.G. and Dagenais, D.L. (1997) Higher Moment Estimators for Linear Regression Models with Errors in the Variables. *Journal of Econometrics*, 76, 193-221.

[10] Diez de los Rios, A. and Garcia, R. (2005) Assessing and Valuing the Nonlinear Structure of Hedge Funds' Returns. Cirano, *Working Paper.*

[11] Doran, H.E. (1992) Constraining Kalman Filter and Smoothing Estimates to Satisfy Time-Varying Restrictions. *Review of Economics and Statistics*, 74, 568-572.

[12] Fama, E.F. and French, K.R. (1993) Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3–56.

[13] Fung, W. and Hsieh, D.A. (1997) Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds. *The Review of Financial Studies*, 10, 275-302.

[14] Fung, W. and Hsieh, D.A. (2001) The Risks in Hedge Fund Strategies: Theory and Evidence From Trend Followers. *Review of Financial Studies*, 14, 313-341.

[15] Fung, W. and Hsieh, D.A. (2002) Risk in Fixed-Income Hedge Fund Styles. *The Journal of Fixed Income*, 12, 6-27.

[16] Getmansky, M., Lo, A.W. and Makarov, I. (2004) An Econometric Model of Serial Correlation and Illiquidity in Hedge Fund Returns. *Journal of Financial Economics*, 74, 529-609.

[17] Glosten, L. and Jagannathan, R. (1994) A Contingent Claim Approach to Performance Evaluation. *Journal of Empirical Finance*, 1, 133-160.

[18] Goldman, M., Sosin, H. and Gatto, M. (1979) Path Dependent Options: 'Buy at the Low, Sell at the High'. *Journal of Finance*, 34, 1111-1127.

[19] Henriksson, R. and Merton R. (1981) On Market Timing and Investment Performance. *Journal of Business*, 54, 513-523.

[20] Hübner, G. and Capocci, D. (2004) Analysis of Hedge Funds Performance. *Journal of Empirical Finance*, 11, 55-89.

[21] Hübner, G. and Papageorgiou, N. (2006) Optional and Conditional Components in Hedge Fund Returns. University of Liège, *Working Paper*.

[22] Kraus A. and Litzenberger R. (1976) Skewness Preference and the Valuation of Risk Assets. *Journal of Finance*, 31, 1085-1100.

[23] Lhabitant, F.-S. (2004) Hedge Funds: Quantitative Insights. *John Wiley & Sons*, 1st edition.

[24] Lobosco, A. and DiBartolomeo, D. (1997) Approximating the Confidence Intervals for Sharpe Style Weights. *Financial Analysts Journal*, 80-85.

[25] Lucas, L. and Riepe, M.W. (1996) The Role of Returns-Based Style Analysis: Understanding, Implementing, and Interpreting the Technique. Ibbotson Associates, *Working Paper*.

[26] Okunev, L. and White, D. (2006) Hedge Fund Risk Factors and the Value at Risk of Fixed Income Trading Strategies. *Journal of Fixed Income*, 16, 46-61.

[27] Samuelson, P.A. (1970) The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances and Higher Moments. *Review of Economics Studies*, 37, 537-542.

[28] Sharpe, W.F. (1988) Determining a Fund's Effective Asset Mix. *Investment Management Review*, 59-69.

[29] Sharpe, W.F. (1992) Asset Allocation: Management Style and Performance Measurement. *Journal of Portfolio Management*, 18, 7-19.

[30] Swinkels, L. and Van Der Sluis, P.J. (2006) Return-Based Style Analysis with Time-Varying Exposures. *European Journal of Finance*, 12, 529-552.

[31] Ter Horst, J.R., Nijman, T.E. and De Roon, F.A. (2004) Evaluating Style Analysis. *Journal of Empirical Finance*, 11, 29-53.