Regular(ized) Hedge Fund Clones*

Daniel Giamouridis[a,1], Sandra Paterlini[b]

[a] Department of Accounting and Finance, Athens University of Economics and Business, Athens, Greece

[b] Department of Economics, University of Modena and Reggio E., Modena, Italy

First draft: October 28, 2008
This draft: January 6, 2008

Abstract

This article addresses the problem of portfolio construction in the context of efficient hedge fund investments replication. We propose a modification to the standard à la Sharpe 'style analysis' where we augment the objective function with a penalty proportional to the sum of the absolute values of the replicating asset weights, i.e. the norm of the asset weights vector. This penalty regularizes the optimization problem, with significant impacts on the stability of the resulting asset mix and the risk and return characteristics of the replicating portfolio. Our results suggest that the norm-constrained replicating portfolios exhibit significant correlations with their benchmarks, often higher than 0.9, have a fraction, i.e. about 1/2 to 2/3, of active positions relative to those determined through the standard method, and are obtained with turnover which is in some instances about 1/4 of that for the standard method. Moreover, the extreme risk of the replicating portfolios obtained through the regularization method is always lower than that exhibited by currently available commercial hedge fund investment replication products.

---

* Daniel Giamouridis is at the Department of Accounting and Finance in the Athens University of Economics and Business, Athens, Greece. He is also Visiting Fellow at Cass Business School, City University, London, UK, and, Research Associate at EDHEC Risk and Asset Management Research Centre, EDHEC Business School, Nice, France. Sandra Paterlini is at the Department of Economics, University of Modena and Reggio E., Modena, Italy. She is also Senior Research Fellow at the Center for Quantitative Risk Analysis, Department of Statistics, LMU Munich, Germany. We would like to thank Alessio Matteuzzi (Citi) for his valuable comments on practical issues. The paper has benefited from the comments of Andrea Resti and one referee, both from Carefin-Bocconi and Lionel Martelini. The paper has also benefited from the comments of participants at the following meetings: University of Minnesota School of Mathematics, Financial Mathematics Seminar Series. Financial support from Carefin-Bocconi Centre for Applied Research in Finance is greatly appreciated. Any remaining errors are our own.
1 Corresponding author. Athens University of Economics and Business, Department of Accounting and Finance, 76 Patission Street, GR-10434 Athens, Greece. Tel: +30 210 8203925; fax: +30 210 8203936.
E-mail address: dgiamour@aueb.gr (D.Giamouridis)

Regular(ized) Hedge Fund Clones

Abstract

This article addresses the problem of portfolio construction in the context of efficient hedge fund investments replication. We propose a modification to the standard à la Sharpe 'style analysis' where we augment the objective function with a penalty proportional to the sum of the absolute values of the replicating asset weights, i.e. the norm of the asset weights vector. This penalty regularizes the optimization problem, with significant impacts on the stability of the resulting asset mix and the risk and return characteristics of the replicating portfolio. Our results suggest that the norm-constrained replicating portfolios exhibit significant correlations with their benchmarks, often higher than 0.9, have a fraction, i.e. about 1/2 to 2/3, of active positions relative to those determined through the standard method, and are obtained with turnover which is in some instances about 1/4 of that for the standard method. Moreover, the extreme risk of the replicating portfolios obtained through the regularization method is always lower than that exhibited by currently available commercial hedge fund investment replication products.

Keywords: Hedge Funds; Replication; Portfolio Management; Index Tracking; Norm-constrained portfolios

JEL Classification: G11; G12; C00

1. Introduction

The last couple of decades have witnessed a rapidly growing interest in hedge funds. Relative to traditional investment portfolios, hedge funds exhibit some unique characteristics: they are flexible with respect to the types of securities they hold and the type of positions they take; they are not subject to public disclosure of their activities; and they are not evaluated against a passive benchmark. This set up encourages hedge fund managers to construct highly dynamic, complex trading strategies and, as a result, expose their portfolios to a plethora of economic risk factors. Not surprisingly the return on alternative investments has shown little correlation with returns on traditional assets such as stocks and bonds. Moreover, in many cases returns on hedge fund investments have also been significantly high[2].

Whether the latter characteristics are due to 'skill' or due to non-conventional techniques such as short selling, leverage and the use of derivatives employed by hedge fund managers has been debated by many researchers. At the core of this strand of literature are studies such as Agarwal and Naik (2000) and Fung et al. (2006) which conclude that the 'alpha'[3] of the average hedge fund or fund of funds manager is very poor and not persistent. Khadani and Lo (2008) in a different context highlight that the fact that the entire class of long/short equity strategies moved together so tightly during August 2007 is consistent with the hypothesis of certain existing common factors within that class. In other words the literature suggests that hedge fund performance – on average – can be attributed primarily to 'alternative betas' rather than 'skill'.

[2] Hedge Fund Research in its 'Second Quarter 2007 Hedge Fund Industry Report' which covers the period January 1997 through to the second quarter of 2007, reports a twelve-month correlation between the overall hedge fund market and S&P 500 of 0.50 on average. The twelve-month correlation between the overall hedge fund market and Lehman Government/Credit Index is at -0.40 on average. Annual Sharpe ratios are 3.58 for the overall hedge fund market and 2.12 and 0.41 for the S&P 500 and Lehman Government/Credit Index respectively.

[3] The component of performance that is attributed to 'skill', i.e. the intercept in the regression of the fund's excess return on the excess return of one or more passive benchmarks.

And this triggers a natural question: Can synthetic hedge funds be created at a lower cost to investors?

The benefits from synthetic hedge funds are several, the most obvious relating to cost, liquidity, transparency, and barriers to entry. A synthetic hedge fund product based on futures, total return swaps, ETFs, or other instruments may involve much lower costs compared to the 2/20 fee structure, i.e. 2% management fee and 20% success fee, which is more or less the hedge fund industry's standard. Moreover, replication through exchange traded assets provides significant liquidity. Hedge funds typically require – at least – one month notice before redemption and many of them have long lock-up periods. In addition, a synthetic structure is transparent as far as the underlying holdings are concerned. Fund investors on the other hand know very little on the allocation of their money, i.e. which assets/trading strategies their money is invested on. Finally, synthetic products may be open for investment to small investors or portfolios wishing to allocate small amounts in the hedge fund universe. Many hedge funds require a minimum investment between $500,000 and $1,000,000 while some funds are not open to new investors.

These observations have attracted the interest of both academics and practitioners and have stimulated a quest to develop effective hedge fund replication strategies. Technically, the development of synthetic hedge fund products can be related to the so-called index tracking problem. Index tracking is a form of passive management where the aim is to reproduce as close as possible a stock market index or the performance of a mutual fund by using a subset of assets or market indexes (see, e.g. Corielli and Marcellino, 2006).

In terms of hedge fund replication the current literature focuses on two different approaches[4]: moment matching and factor based replication. Moment matching has been introduced and thoroughly studied by Kat and Palaro (2005a, b) and used by Amenc et al. (2007), and Papageorgiou et al. (2008). The core idea is developed in Amin and Kat (2003) in the context of hedge fund performance measurement. Central to this approach is the conjecture that the return profile of an investment strategy can be replicated if its risk profile, i.e. volatility, skewness, kurtosis, correlation to a specified portfolio, are matched. The replicating strategy involves holdings on: a) a reserve portfolio - the return driver - which is always a long position, and b) the investor's portfolio, which is the reference for the correlation target with a potential short position. While favourable empirical properties of this approach are acknowledged in the literature several criticisms are put forward (see, e.g. Amenc et al., 2007). The most important among them is that the probability distributions change and the replicating portfolio may not be able to track those changes. It is also argued that that the first moment may not match automatically when matching the risk profile.

Factor based replication has so far gained significant consensus and is currently to our knowledge the most widely applied methodology in the practice of hedge fund replication[5]. Hasanhodzic and Lo (2007) provide a detailed overview of this approach as well as the results of a very comprehensive empirical analysis. Basically, factor based replication builds upon Sharpe's

[4] One third approach termed 'security based replication' is based on the implementation of a generic version of a given strategy. Fung and Hsieh (2001) for example use lookback straddles to model trend following strategies which are typical for CTAs. Schneeweis and Kazemi (2007) take long/short positions in commodity and financial futures using moving average signals to replicate trend-following CTAs. Mitchell and Pulvino (2001) study the risk/return profile of the merger arbitrage strategy through long portfolios of target firms and short positions in acquiring firms. Convertible Arbitrage is replicated through long positions in convertible bonds and short positions in equities in Agarwal et al. (2006) while fixed income arbitrage is studied in Duarte et al. (2007). Security based replication however may not be that different from a typical hedge fund.

[5] Meryll Lynch's Equity Volatility Arbitrage Index and Factor Index for example build on the notion of factor based replication and so do Deutsche Bank's Absolute Return Beta Index, Goldman Sachs' Absolute Return Tracker Index as well as Credit Suisse's recent alternative beta initiative.

(1992) style analysis, i.e. a constrained beta linear regression on a given set of factors to build the best mimicking portfolio. Hasanhodzic and Lo (2007) show that this approach works well within the Equity Market Neutral, Global Macro, Long/Short Equity Hedge, Managed Futures, Multi-Strategy, and Fund of Funds space. Replication is not however successful for Event Driven and Emerging Markets funds. Jeager and Wagner (2005), Amenc et al. (2007), and Amenc et al. (2008) provide additional evidence for the factor based approach.

Despite the attractiveness of factor-based hedge fund replication, it has not yet to our knowledge considered important empirical conclusions from the hedge fund risk literature. Many studies (see, e.g. Agarwal and Naik, 2004; Fung and Hsieh, 2000, 2001; Liang, 1999, Vrontos et al. 2008) argue that the true set of hedge fund risk factors is virtually unknown – due to the lack of transparency and the large number of possible market and trading strategy combinations a manager can follow – and use statistical techniques to identify important factors. Factor-based hedge fund replication has thus far considered a fixed set of factors, typically a small set, and relied on optimization to determine the amount by which an asset contributes to the mimicking portfolio. As we show in our empirical analysis this approach tends to allocate capital to all assets in the original universe thus failing to distinguish in an efficient way between important and less important hedge fund return drivers. This is an extremely critical issue when the possible number of factors is large.

In addition, the constrained regression seems to have been applied in a rather traditional way without accounting for the criticisms and remedies related to such problems. As we show in our empirical analysis standard constrained optimization of the weights in the mimicking portfolio results in a very unstable mix over time. Moreover many of the assets in the replicating portfolio possess extreme weights. Finally, current studies have to our knowledge not looked at the properties of the mimicking portfolio in detail. None of the studies we know are concerned with the cost of implementation of the replicating strategy or the extreme risk of the mimicking portfolio for example. After all, setting up a hedge fund clone investment programme involves tactical trading of the underlying factors as well as continuous monitoring and management of its risk.

To fill the gaps outlined above, we propose a new methodology for the construction of hedge fund clones. Our approach relies on minimizing a penalized version of the tracking error volatility between the hedge fund index and the clone. The penalty we consider is proportional to the sum of the absolute values of the replicating asset weights, i.e. the norm of the asset weights vector. This penalty regularizes the optimization problem and jointly addresses model selection and estimation error. In the context of portfolio construction – a problem that is analogous to the construction of mimicking portfolios – norm-constrained portfolios have shown to possess superior properties compared to those obtained from conventional approaches but also compared to those obtained from a number of alternative improved strategies proposed in the literature (see, e.g. Brodie et al. 2007, DeMiguel et al., 2007). It can be shown that our approach is a special case of the standard à la Sharpe 'style analysis'. In our sample the factor-based hedge fund clones are extremely stable in terms of the exposures to systematic risks and possess improved risk and return characteristics. In particular, the norm-constrained replicating portfolios exhibit significant correlations with their benchmarks, often higher than 0.9, have a fraction, i.e. about 1/2 to 2/3, of active positions relative to those determined through the standard method, and are obtained with turnover which is in some instances about 1/4 of that for the standard method. Moreover, the extreme risk of the replicating portfolios obtained through the regularization method is always lower than that exhibited by currently available commercial hedge fund investment replication products.

The contributions of this article are several. First, we propose a method that has been successfully applied in portfolio construction studies for the construction of hedge fund clones. Second we study for the first time the properties of the clones in a broader framework which extends to the practical implementation of the developed strategy. Third we study the out-of-sample properties of the clone from the perspective of a structured products manager who has not only to create the clone but also to manage its risks as well as the transaction costs of the replicating strategy. Fourth we provide additional evidence for the use of regularization methods in empirical finance.

The paper is organized as follows. In Section 2 we describe the proposed methodology that is used to determine the synthesis of the replicating portfolio. Section 3 discusses the data and the structure of our empirical experiments. In Section 4 we present the results of the model selection analysis. Section 5 presents the results from the out-of-sample analysis of the properties of the mimicking portfolios as well as the robustness checks for our results and Section 6 concludes.

2. Methodologies

Our approach falls in the class of factor based replication methods. As in Sharpe (1992) and Hasanhodzic and Lo (2007) among others, we use a decomposition of the return time series $\mathbf{r}_i$ of a hedge fund index into several components factors and an idiosyncratic component $\boldsymbol{\varepsilon}$, according to:

$$\mathbf{r} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad (1)$$

In Equation (1), $\mathbf{r} = (r_t,...,r_T)$ is the $(T\text{-}t\text{+}1)$x1 vector of a hedge fund index return time series, $\mathbf{F}$ is the $(T\text{-}t\text{+}1)$xN matrix whose k^{th} ($k = t,\ldots,T$) row is the 1xN vector $\mathbf{f}_t$ of returns of the N factors at time k, i.e. $\mathbf{F} = (\mathbf{f}_k)$, $\boldsymbol{\beta} = (\beta_1,..\beta_N)^T$ is the Nx1 vector of factor loadings in the time period [t,T] and $\boldsymbol{\varepsilon}=(\varepsilon_t,...,\varepsilon_T)$ is the $(T\text{-}t\text{+}1)$x1 vector of the idiosyncratic component.

To construct a mimicking portfolio for a hedge fund index with increased tracking ability it is necessary to identify the appropriate replicating factors first. The standard approach involves a fixed set of factors which are assumed a priori to be appropriate for replicating the returns of hedge fund index returns. Hasanhodzic and Lo (2007) for example undertake their analysis with five factors, i.e. proxies for the equity, bond, currency, credit, and commodity markets. Alternatively, the hedge fund risk literature has suggested that typical model selection approaches such as stepwise regression, selection based on Akaike's (1973) information criterion or the Schwarz's (1978) Bayesian information criterion (see, e.g. Agarwal and Naik, 2004, Vrontos et al. 2008) may be used to identify appropriate factors. With the chosen set of factors, the composition of the replicating portfolios can be determined through a constrained regression, i.e. beta coefficients are estimated with the constraint that they sum up to one, which can be expressed as:

$$\hat{\boldsymbol{\beta}}^{Sharpe} = \underset{\boldsymbol{\beta}}{\text{minimize}} \left\| \mathbf{r} - \mathbf{F}\boldsymbol{\beta} \right\|_2^2 \qquad (2)$$
$$\text{s.t. } \mathbf{1}_N \boldsymbol{\beta} = 1$$

where $\| \mathbf{r} - \mathbf{F}\boldsymbol{\beta} \|_2^2$ is the 2-norm of vector $\left(\mathbf{r} - \mathbf{F}\boldsymbol{\beta}\right)$, i.e. $\| \mathbf{r} - \mathbf{F}\boldsymbol{\beta} \|_2^2 = \sum_{t=1}^{T} \left(\mathbf{r} - \mathbf{F}\boldsymbol{\beta}\right)_t^2$, the sum of squared residuals (SSR hereafter) or tracking error variance, and $\mathbf{1}_N$ is a 1xN vector of ones. The elements of $\hat{\boldsymbol{\beta}}^{Sharpe}$ are then used as portfolio weights for the chosen factors, hence the replicating portfolio returns are equivalent to the fitted values $\mathbf{r}$.

Brodie et al. (2007) highlight that in the financial context it is likely that a standard numerical procedure for the optimization in Equation (2) may amplify the effects of noise anisotropically, leading to an unstable and unreliable estimate of the vector **β**. Moreover, with little ex ante knowledge regarding the exact factors that drive hedge fund index returns it seems appropriate that a broad set is initially considered. This renders the optimization problem more computationally complex and increases the possibility that unstable estimates of the vector **β** are obtained. A comprehensive set of replicating factors may also involve collinear variables – even just during certain periods, i.e. in distress markets conditions – which entail additional obscurity for the numerical procedure (see, Brodie et al., 2007).

To tackle such issues, recent works in the portfolio construction literature (see, e.g. Brodie et al., 2007, De Miguel et al. 2007, Lobo et al. (in press), Welsch and Zhou, 2007) have proposed the use of regularization procedures such as ridge regression (Hoerl and Kennard, 1970) or lasso regression (Tibshirani, 1996). The evidence so far suggests that such methods reduce the sensitivity of the optimization to possible collinearities between assets, control transaction costs by promoting sparsity[6], and improve the out-of-sample performance relative to the classical Markowitz minimum variance portfolio. Furthermore, regularization methods have shown to possess interesting Bayesian and moment-shrinkage interpretations (Tibshirani, 1996, De Miguel et al., 2007).

Despite these appealing properties and the analogy of the portfolio construction and index tracking problems, regularization methods have not to our knowledge yet been investigated in the context of hedge fund cloning or index tracking in general. Our investigation is based on both ridge and lasso regression.

[6] A sparse model is a model where very few assets are selected.

Ridge regression involves the optimization of a penalized version of the SSR function. The penalty term is proportional to the 2-norm of the weights. Ignoring optimization constraints for ease of exposition, we can express the ridge regression method as:

$$\hat{\boldsymbol{\beta}}^{Ridge} = \arg\min_{\boldsymbol{\beta}} \left[\left\| \mathbf{r} - \mathbf{F}\boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\beta} \right\|_2^2 \right] \qquad (3)$$

We denote by $\| \mathbf{x} \|_p^p$ the sum $\sum_{t=1}^{T} | x_t |^p$, i.e. the p-norm of vector $\mathbf{x}$. Adding the penalty in the objective function results in adding a positive constant to the diagonals of $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ such that the matrix $\mathbf{F}^{\mathrm{T}}\mathbf{F}$ is non-singular and computing the inverse does not affect the stability of the SSR minimization problem (see Appendix). Daubechies et al. (2004) have shown that any p-norm penalty ($1 \le p \le 2$) suffices to stabilize the minimization of Equation (2) by regularizing the inverse problem. When p=2, the optimization problem is still a continuously differentiable optimization problem and it can be easily tackled by conventional techniques, while providing more stable and accurate estimates than standard SSR minimization. However, with p=2 the coefficients although shrunk – thus stable – cannot become 0. Hence, the resulting replicating portfolio will not be parsimonious.

To accommodate this possibility yet sustain useful properties of regularization methods, Tibshirani (1996) proposed to add a penalty proportional to the 1-norm, an approach termed lasso regression. Ignoring optimization constraints for ease of exposition, we can express the lasso regression method as:

$$\hat{\boldsymbol{\beta}}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \left[\left\| \mathbf{r} - \mathbf{F}\boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\beta} \right\|_1^1 \right] \qquad (4a)$$

It can be easily shown (see Brodie et al., 2007) that this is equivalent to solving the following optimization problem:

$$\hat{\boldsymbol{\beta}}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \|\mathbf{r} - \mathbf{F}\boldsymbol{\beta}\|_2^2 \qquad \text{(4b)}$$

$$\text{s.t. } \|\boldsymbol{\beta}\|_1^1 \leq t$$

Due to the geometrical nature of such constraints, the resulting procedure can give coefficient estimates that are exactly zero in the cases when the corresponding factors contain little information about the hedge fund index (see Appendix). Thus, lasso regression involves model selection and estimation in a single step. Brodie et al. (2007) list several additional useful consequences of the regularization with the 1-norm penalty including stability, control of the short positions, and account of transaction costs. In particular, under the budget constraint the λ penalty becomes a penalty on short positions. The larger the λ, the smaller the number of factors to be selected and the larger the number of no-short-positions in the replicating portfolio. Transaction costs on the other hand can be controlled for through a modified penalty term, i.e. $\sum_{k=1}^{N} s_k |\beta_k|$, where s_k is the bid-ask spread for the k^{th} security.

The choice of λ in ridge and lasso regression is critical to control several aspects of the optimization problem. Although it is arbitrarily chosen in some instances (see, e.g. Brodie et al., 2007), more sophisticated cross-validation methodologies based on statistical or economic criteria could be implemented to determine its value (see, e.g. De Miguel et al., 2007).

To construct a linear clone for a hedge fund index we impose the additional constraint that asset weights vary between -1 and 1[7]. So in the case of lasso regression, the method we propose as most appropriate for hedge fund index replication, the optimization problem becomes:

$$\hat{\boldsymbol{\beta}}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \left[\left\| \mathbf{r} - \mathbf{F}\boldsymbol{\beta} \right\|_2^2 + \lambda \left\| \boldsymbol{\beta} \right\|_1^1 \right] \qquad (5)$$

s.t.

(a) $\mathbf{1}_N \boldsymbol{\beta} = 1$

(b) $-1 \leq \boldsymbol{\beta} \leq 1$

When using our approach, constraint (b) is not necessary to get realistic asset weights since increasing the value of λ, results in a decrease in the number of short positions and shrinks the portfolio weights. However, we introduce this constraint to have a more fair comparison with a standard quadratic programming approach[8]. This problem nests Sharpe's (1992) 'style analysis'. In fact, 'strong style analysis'[9] constrains the portfolio weights to be positive and to add to one. This is equivalent to constraining the 1-norm of the factor loadings to be equal to 1. Hence, it is not surprising that imposing no short-selling constraints on portfolio weights as shown in Jagannathan and Ma (1993) has a shrinking effect which provides better prediction accuracy.

---

[7] $\boldsymbol{\beta}$ may in practice vary depending on the manager's mandate. Given that the replicating portfolio is by construction a low volatility portfolio it should involve either low beta exposures or high beta exposures that net each other out. High values of beta however should be avoided in principle for two main reasons. First, traders are averse to holding highly leveraged positions in single assets for risk purposes. And second, implementing strategies involving high betas encompasses high transaction costs associate with the unwinding/creation of large positions. From a methodological point, allowing the weights to vary in [-1, 1] allows enough space to investigate the stability and assess the feasibility of the alternative methodologies. Hence, our choice of $\boldsymbol{\beta}$ constraints while by no means favour the proposed approach serve as reasonable choice from a practical and methodological standpoint.

[8] Adding constraints (a) and (b) can affect the choice of λ in order to enforce sparsity. In fact, if no constraints are imposed on $\boldsymbol{\beta}$, the 1-norm can vary between 0 and infinity, while if we impose constraints (a) and (b) the 1-norm is bounded between 1 and N-1. Hence, larger λ values are required when imposing such constraint to enforce sparsity into the model.

[9] Depending on the constraints on the portfolio weights, 'style analysis' can be classified as 'weak', i.e. no constraints are imposed on the portfolio weights, 'semi-strong', i.e. only the budget constraint is considered (the sum of portfolio weights equals 1), and 'strong', i.e. portfolio weights are required to be positive and sum up to 1 (see, e.g. Horst et al., 2004).

3. Data and empirical design

We carry out our analysis with monthly hedge fund index return data from the Hedge Fund Research (HFR hereafter) database. Prior studies that focus on hedge fund index replication include Amenc et al. (2007), Amenc et al. (2008). Other studies focus on the replication of individual funds, see for example Hasanhodzic and Lo (2007), Jeager and Wagner (2005).

Our dataset covers the period February 1990 to July 2008. We consider the four primary strategy indices namely Equity Hedge, Event Driven, Macro, and Relative Value. Equity Hedge comprises managers that maintain positions both long and short in primarily equity and equity derivative securities. Even Driven classified funds are those investing on companies that are likely to be involved in corporate transactions of a wide variety. The funds constituting the Macro index engage in a broad range of strategies in which the investment process builds upon predictions of movements in underlying economic variables and the impact these have on equity, fixed income, and currency and commodity markets. Relative Value managers speculate on realization of a valuation discrepancy in the relationship between multiple securities. In addition to the primary strategy indices we consider an index with a regional focus, namely the Emerging Markets index. This index comprises managers independent of their investment strategy. As a proxy for the aggregate market we consider the Fund Weighted Composite Index which is an equal-weighted index made of over 2,000 individual funds. Finally, we consider a proxy for the overall Fund of Funds market, the Fund of Funds Composite Index which comprises over 800 individual Fund of Funds based on equal-weighting.

We investigate if these indices can be explained by common risk factors representing almost all traditional asset classes, i.e. equity, bond, commodity, currency, and real estate. Within each asset class we consider various detailed investment styles to expand the cross-section of risk exposures for the typical hedge fund, i.e. large cap stocks, volatility, emerging markets, government bonds, credit, oil, gold among others. We insure that each of the factor returns can be realized through relatively liquid instruments, and therefore the returns of the mimicking portfolio may be achievable, by considering indices with active exchange traded or over-the-counter derivatives markets. Table 1 lists the indices that we consider as replicating factors in our analysis. Our approach seeks to investigate the added value of model selection techniques in the context of hedge fund cloning hence the set of possible risk factors is very comprehensive[10]. This choice is motivated by the conclusions in several studies highlighting that model selection may have important implications for style analysis (see, e.g. Dupleich et al., 2008), performance attribution (see, e.g. Agarwal and Naik, 2004, Vrontos et al. 2008), and hedge fund return predictability (see, e.g. Amenc et al., 2003, Vrontos and Giamouridis, 2008). Data for the replicating factors are obtained through Bloomberg for the period February 1990 to July 2008. In the full sample the correlations between factors range from -0.67 to about 1. The average correlation between the factors is 0.14 with standard deviation 0.36. Correlations are distributed with positive skewness, i.e. 0.62, while the first quartile is at -0.08 and the third 0.41. We choose not to remove factors that are highly correlated to be able to investigate the impact this situation may have on the model selection techniques.

Insert Table 1 somewhere here

[10] Prior studies have concluded that hedge fund returns have non-linear patterns and have proposed non-linear risk factors to explain their returns. Agarwal and Naik (2004) for example show that hedge funds are exposed to a factor that captures the return characteristics of an out-of-the-money put option, Agarwal, et al. (2008) find that skewness and kurtosis factors obtained through options' cross-sections are important variables in hedge fund risk modelling, while Fung and Hsieh (2001) show that trend following strategies are explained by lookback straddles. While these factors may be important for risk management purposes, their replication involves trading instruments that may in some instances be beyond the mandate of the manager, e.g. over-the-counter lookback straddles. To maintain generality and also since our primary purpose is to investigate the relative performance of the proposed methodology we focus on a broad set of linear factors motivated by current academic and practitioner work on hedge fund return replication.

The empirical analysis is undertaken on a rolling-window basis. We initially consider a rolling-window of 120 months which is then carried forward by one month as the window moves through the data set. Each month we seek the best model through five different model selection strategies: (a) straight multiple regression (Straight hereafter), i.e. estimation of Equation (1), where the best model is that containing the significant factors, (b) factor selection based on Akaike's (1973) Information Criterion (AIC hereafter), (c) factor selection based on Schwarz's (1978) Bayesian Information Criterion (BIC hereafter), (d) ridge regression (Ridge hereafter) where the best model is the one determined by solving the optimization problem in Equation (3), and (e) lasso regression (Lasso hereafter) where the best model is the one determined by solving the optimization problem in Equation (4a), with two different values for λ. For AIC and BIC we estimate all possible models with respect to Equation (1), i.e. 2^K models arising from all possible combinations of the K factors. We identify the best model as that with the lowest value of the respective information criterion, thus we jointly take into account model fit and complexity.

4. Factor exposures

To determine the explanatory power of the factors in our set, we perform a time-series analysis of each hedge fund index in our sample. In particular, within a rolling-window set-up we apply the model selection strategies (a) to (e) detailed in the Section 3. We look into various aspects of the estimated models that are very important for hedge fund investment replication. First we discuss the nature of the important factors. Next, we examine the in- and out-of-sample fit of the estimated models. We also report the size of the selected model, i.e. the number of important factors. And finally we discuss the maximum and minimum factor exposures.

Table 2 reports the factors that are identified by the respective model selection strategies as valuable explanatory variables. The straight multiple regression and the ridge regression indicate that all factors are important, i.e. the coefficients in Equations (1) and (3) are not exactly equal to zero, therefore we report results only for AIC, BIC, and Lasso. Overall, the results indicate that hedge fund portfolios involve holdings in literally all traditional asset classes. Equities, bonds, commodities, currencies, and real estate assets appear as important drivers for the various primary strategy managers as well as for the emerging markets manager, the aggregate hedge fund manager, and the aggregate fund of funds manager. Among the most important variables are Russell 2000 Total Return Index and the MSCI Emerging Markets Index suggesting significant exposures to the US and Emerging Markets equity space. A significant exposure to the dollar index could be motivated by the substantial activity in carry-trading as well as cross-currency investments. In terms of commodity investments managers seem to have favoured oil and gold over the study period while the rest of commodities represented by the S&P GSCI Index seems to have also played an important role. Cash and government bonds are important factors for many strategies while investment-grade bonds are related primarily to event-driven strategies. Interestingly, Europe, Australasia, and Far East equity markets have not been among the top investment venues.

Insert Table 2 somewhere here

The fit of the estimated models is examined through the results presented in Table 3 where we tabulate the in- and out-of-sample Root Mean Squared Error (RMSE hereafter). The figures in Table 3 are annualized RMSE computed for the best model in each model selection strategy over the entire period. The minimum in-sample RMSE, i.e. 2.39%, is obtained for ‘Straight’ for the Fund Weighted Composite Index. The maximum in-sample RMSE, i.e. 7.41%, is observed for ‘BIC’ and ‘Lasso (high)’ for the Emerging Markets index. Out-of-sample RMSE is lowest for ‘Lasso (low)’ for the Relative Value managers and highest for ‘Straight’ applied to Emerging

Markets managers, 2.49% and 5.85% respectively. These figures are lower compared to those reported in Amenc et al. (2008) who consider linear and the non-linear models to replicate similar indices[11]. Overall, we conclude that in terms of 'average' fit both in- and out-of-sample neither of the methodologies seem to compare favourably to the others. Ridge regression has a slightly better out-of sample performance, which seems to confirm the need to use penalized least squared errors approaches when dealing with financial factor models. A noticeable property for 'Lasso' however is that in most cases determines models with better out-of-sample fit relative to their in-sample fit. This result provides additional support to DeMiguel et al. (2007) who argue that 'Lasso' can be very successful out-of-sample in the presence of estimation error.

Insert Table 3 somewhere here

Next, we turn our eye to examine the size of the best model obtained. While we have already highlighted that 'Straight' and 'Ridge' result in non-zero factor exposures for all factors, 'AIC', 'BIC', and 'Lasso' penalize large models. In Table 4 we report the average number of factors in the best models selected through 'AIC', 'BIC', and 'Lasso'. We conclude that models obtained through 'AIC' and 'Lasso (low)' are generally large. 'BIC' and 'Lasso (high)' determine models that are on average of similar size, although models obtained through 'BIC' are almost always smaller. Examining these results jointly with those discussed above (see Table 3), we conclude that it is possible to achieve a significant degree of model fit by focusing on a subset of factors – obtained through 'AIC', 'BIC', and 'Lasso' – which is on average half the size of the full set. In particular, 'BIC' and 'Lasso (high)' seem to offer a unique trade-off between parsimony and out-of-sample fit. The best models selected through 'BIC' and 'Lasso (high)' are the smallest in size

[11] This is a rough comparison given that Amenc et al. (2008) are studying a different period, i.e. January 1997 to December 2006, and a different database, i.e. TASS. Since we do not test the effect of differences in the two sample we treat this comparison as indicative.

while at the same time they provide comparable out-of-sample fit with the other model selection techniques.

Insert Table 4 somewhere here

One final aspect we examine is the stability of the estimated parameters, rather the tendency of the model selection techniques to result in models with extreme factor exposures. In Table 5 we summarize the results of this analysis. We show the minimum and maximum exposures on the factors obtained through the alternative model selection techniques over the entire period. We observe incredibly low exposures, i.e. the minimum exposure is -10.03, as well as unreasonably high exposures, i.e. the maximum exposure is 10.03, are obtained with 'Straight'. 'Ridge', 'AIC' and 'BIC' compute exposures not as extreme as those in 'Straight', still far smaller than -1.00 and far bigger than 1.00 in many instances. Interestingly, the picture is very different for 'Lasso'. For both values low and high values of λ the minimum factor exposures are close to zero, the minimum of them being -0.20. The maximum factor exposures are not extreme either with the maximum of them being 1.69.

Insert Table 5 somewhere here

Overall, in this section we show that rigorous model selection techniques i.e. 'AIC', 'BIC', and 'Lasso' can be used to obtain reduced models, with economically relevant factors, without trading off fit relative to a model containing a broader set of factors. Moreover we show that 'Lasso' can be very successful out-of-sample. However 'AIC' and 'Lasso (low)' may result in models whose size is significantly larger than the size of models obtained through 'BIC' or 'Lasso (high)'. When we examine the magnitude of the extreme factor exposures we conclude that 'AIC' and 'BIC' identify factor with extremely high or extremely low exposures.

Collectively, we find that 'Lasso' improves our ability to select important economically relevant factors without trading off fit or parsimony.

5. Clones performance

The analysis in Section 4 reveals that 'Lasso' posses some very attractive properties that can make it a superior method in the context of hedge fund return replication. While we impose no constraints on the weights in the preceding analysis, we can draw important inferences for the core problem at hand – the construction of a hedge fund mimicking portfolio. We show that 'Lasso' can achieve very good out-of-sample fit which is critical for the assessment of the replicating portfolio. We also show that the size of the 'Lasso' best model, and therefore the number of assets in the mimicking portfolio is reduced compared to the size of the portfolio obtained from conventional techniques. This is a very important aspect in the context of hedge fund replication as it is preferable to trade fewer assets in the replicating investment programme. Finally, we show that the extreme factor exposures are within reasonable bounds in 'Lasso' relative to all other approaches. As a result the weights of the different assets in the mimicking portfolio are more likely to be stable over time, consistent with low transaction costs and ease of trading.

This Section investigates whether these properties can be exploited in practice. In particular we consider the portfolio construction problem described in Equation (5) which involves some basic constraints that are typically used in the construction of hedge fund replicating products. We also use a modified framework of Sharpe's (1992) 'style analysis', i.e. the optimization problem described by Equation (2) with the additional constraint that $-1 \leq \boldsymbol{\beta} \leq 1$ [12], as a benchmark approach. We explore several properties of the replicating portfolio including out-of-sample

[12] See also footnote 7.

correlation with the replicated index, the number of asset in the replicating portfolio, performance measures such as the excess return, the tracking error volatility, but also the turnover of the mimicking portfolio. In addition we compare the clones produced with our analysis with selected currently available hedge fund replication products. And finally, we carry out a set of tests to check the robustness of our results.

5.1 Comparison with the benchmark

First we examine the diversity of the replicating portfolio by looking at the number of assets that are identified as important in replicating the benchmark returns. We also examine the out-of-sample correlation of the replicating portfolio returns with the respective hedge fund indices. We report the results of this analysis in Table 6. These results suggest that the mix of assets obtained with 'Lasso' is reduced in size compared to the mix obtained through the standard approach. Depending on the considered values of λ, 'Lasso' can result to a reduced size in the range of about 50% to 75% of the full set [13]. On an absolute basis we observe that both methods are capable of generating portfolio mixes that in most instances have significant correlation with the benchmark indices – the correlation for Equity Hedge, Emerging Markets, and the Fund Composite Index exceed 0.90. The correlation for Event Driven and Fund of Funds managers is about 0.80. Macro and Relative Value managers mimicking portfolios have lower correlation 0.62 and 0.43 with their benchmarks respectively.

Insert Table 6 somewhere here

Next we examine several metrics related to the performance of the replicating portfolio with respect to its benchmark. We compute the average excess return, i.e. the difference between the replicating portfolio return and the return of the benchmark, the tracking error volatility, and the

[13] With Lasso, depending on the choice of λ, we can determine models with a number of factors varying from 1 to N.

turnover that is required by the rebalancing strategy. We report these results in Table 7. We observe that the constrained regression method produces both positive and negative excess returns while the excess return of portfolios constructed through 'Lasso' is always negative on average. When we compare the excess returns of the portfolio obtained with constrained regression and the returns of the 'Lasso' portfolios, we cannot reject the hypothesis that their means are equal, at the 5% significance level. When we compare these results with the results reported in Amenc et al. (2008) for their best performing method, i.e. Kalman filter approach, we conclude that 'Lasso' improves our ability to construct hedge fund replicating portfolios. For example for the Fund of Hedge Funds and the Event Driven indices the average excess return reported in Amenc et al. (2008) is about -0.33% which is worse than -0.13% and almost identical to -0.32% for the respective indices in our experiments for 'Lasso (high)'. For Emerging Markets and Macro Amenc et al. (2008) report -0.79% and -0.67% respectively while our analysis yields -0.49% and -0.23%. In terms of tracking error volatility 'Lasso' does as good and some times even better that the constrained regression approach. One interesting property of the replicating portfolio rebalancing strategy is the turnover which obviously relates to the transaction costs required for obtaining the targeted return. We report the turnover of the portfolio strategy in the last columns of Table 7. The results are consistent with the argument that 'Lasso' allows implicitly to control for turnover. Compared with the constrained regression approach which on average requires about 1/3 of the portfolio to rebalance at each point of time, 'Lasso' portfolios drop that to about 1/15 for the high λ case. Transaction costs for the assets considered in our mimicking portfolio may vary in the range of 5 to 30 bps[14]. Given that excess returns are not statistically different, the benefits in the return can be in the magnitude of 1 to about 10 bps per month. The benefits can be even higher when more frequent rebalancing is necessary, i.e. when

[14] This was an indication provided by practitioners trading such contracts.

structured products on the index should be managed. This issue has not been discussed in previous works to our knowledge although we believe is very critical for the success of the replication strategy.

Insert Table 7 somewhere here

Another aspect that has not been addressed in the current literature relates to the concentration of a large proportion of the replicating portfolio capital on certain assets. In Table 8 we present evidence on the replicating portfolio high weights, i.e. greater than 50%, either long or short. In column headed 'no factors' we present the number of factors which had a high loading during the sample period. In column headed 'average' we present what proportion of the test period on average have the loadings on these factors been high. The column head 'max' presents the max proportion of the test period that a factor has exceeded the threshold. The results in Table 8 indicate that with conventional portfolio optimization a large number of factors concentrate significantly high weights which sometimes are kept at high levels for the entire period. When we consider the results reported for 'Lasso' we observe that the number of factors that a large proportion of wealth is allocated to is significantly lower, for example 0 to 2 factors for 'Lasso (high)' and shows very little persistence.

Insert Table 8 somewhere here

In summary, the results in this section show that norm-constrained hedge fund replicating portfolios are relatively small portfolios in terms of the number of assets they involve and have similar correlation properties with the benchmark indices with portfolios constructed with conventional methods. While their excess return is slightly lower - although not different statistically - the turnover is a fraction of that of a conventionally constructed mimicking portfolio. Finally, although they comprise a significantly smaller number of assets, norm-constrained hedge fund replicating portfolios encompass more evenly weighted mixes of assets.

5.2 Comparison with selected commercially available hedge fund clones

This section discusses the properties of replicating portfolios constructed with 'Lasso' and conventional methods with the properties of selected commercially available hedge fund clones that banks trade with their clients as plain vanilla notes or in the form of structured products. Our analysis involves Citi's HARP Index, Deutsche Bank's Absolute Return Beta Index, and Meryll Lynch's Factor Index. Citi's HARP Index and Deutsche Bank's Absolute Return Beta Index are benchmarked against the HFR Fund of Hedge Funds Index, while Meryll Lynch's Factor Index benchmark is the HFR Fund Weighted Composite index. Data on these indices are obtained from Bloomberg. A word of caution here is due for the fact that the return on the replicating strategies is gross of transaction costs or any other fees.

Table 9 tabulates the results of this analysis. In particular we report risk and return characteristics. We also report extreme risk measures, i.e. Value at Risk and Conditional Value at Risk. When we compare the proposed portfolio construction strategy benchmarked at the Fund Weighted Composite Index with Meryll Lynch's Factor Index we observe several important improvements. Not only the average return is higher, closer to the benchmark, but also the return/risk ratio is higher. Most importantly Meryll Lynch's Factor Index seems to exhibit significant tail risk given the large value of VaR and CVaR at the 1% level, i.e. 5.20% and 5.64% respectively, compared to 3.17% and 3.27% for 'Lasso (low)', and to 3.04% and 3.15% for 'Lasso (high)'. However we need to acknowledge that Meryll Lynch's Factor Index involves six factors whereas 'Lasso' about eleven to fourteen on average with very low turnover however. For the products related to the Fund of Hedge Funds benchmark we observe that the average return of Citi's HARP Index and Deutsche Bank's Absolute Return Beta Index are higher than 'Lasso'. Their volatility is such though that the return/risk ratio is higher for 'Lasso'. The results on the extreme risk measures suggest that the tail risk associated with Citi's HARP Index and Deutsche Bank's Absolute Return Beta Index is higher than the extreme risk of the 'Lasso' portfolio – especially for Deutsche Bank's Absolute Return Beta.

Insert Table 9 somewhere here

Overall the results in this section indicate that 'Lasso' provides an attractive alternative to the methodologies that appear to underlie selected commercially available hedge fund clones, especially from a risk management perspective.

5.3 Robustness checks and additional results

In a separate work (available in detail from the authors upon request) we have extended our analysis to incorporate: (a) a variety of hedge fund indices, (b) additional optimization constraints, (c) a range of values for λ, (d) different estimation windows.

We used a wide range of sub-strategy indices. These included: Equity Market Neutral, Quantitative Directional, and Short Bias form the Equity Hedge universe; Distressed/Restructuring and Merger Arbitrage managers from the Event-Driven universe; Systematic Diversified from the Macro primary strategy; Fixed Income-Convertible Arbitrage, Fixed Income-Corporate, and Multi-Strategy managers from the Relative Value universe; Asia ex-Japan from Emerging Markets; and Conservative, Diversified, Market Defensive, and Strategic managers from the Fund of Funds universe. The results we obtained were qualitatively similar with those we report in the previous sections.

We also tested a variety of optimization constraints. One of the main tests we carried out in this context was to explore situations where leverage is allowed. We relaxed the assumption that $\sum_{i=1}^{N} \beta_i = 1$ to allow for total wealth leveraged by a factor of 1.5, i.e. $\sum_{i=1}^{N} \beta_i = 1.5$. Our approach for incorporating leverage included allocating the differential capital of our allocations to the risk free asset[15]. The results of this analysis showed that the performance of the replicating portfolio does not improve. This in fact is reasonable as requiring that leverage is 150% both in good and bad times may not be a wise choice in practice. An improvement to this may be to set a cap on leverage and/or link allowed leverage with market conditions or momentum. We also varied the lower and upper bound constraints for β_i. While this had a considerable impact on the clones determined through constrained regression it changed the behaviour of the mimicking portfolios obtained through 'Lasso' only marginally.

The results reported in the previous Sections are obtained for arbitrarily values of λ. In particular in Section 5 (constrained optimization) values are such that for low (high) λ about 75% (60%) of the total number factors are being picked up. In Section 4 (unconstrained optimization) a high λ that is four times the value of low λ identifies models with on average half the number of factors. We tested additional values for λ. We observed that if we keep λ too high we tend to overfit the model in-sample (we do slightly better than with a lower penalty), but then out-of sample the performance becomes worse. Furthermore, we investigated standard statistical cross-validation

[15] In particular we estimated the weights in the replicating portfolio through $r = \sum_{i=1}^{K} \beta_i F_i + \varepsilon$ and computed the return of the final portfolio through $r = \sum_{i=1}^{K} \beta_i F_i + r_f \times \left(1 - \sum_{i=1}^{K} \beta_i\right)$, where r_f is the risk-free rate. Alternatively one can estimate the weights through the equation $r - r_f = \sum_{i=1}^{K} \beta_i \left(r_i - r_f\right) + \varepsilon$.

methods. These, did not in fact provide better results in term of accuracy, while resulted in higher turnover. Future research may focus on different criteria for choosing λ.

In terms of the estimation window, our robustness analysis concluded that with shorter estimation periods, i.e. a window of 60 months. The analysis showed that with shorter estimation windows the weights become less stable. In relative terms however the impact is considerably stronger for constrained regression mimicking portfolios than it is for portfolios obtained through 'Lasso'.

In summary, additional analysis concludes that the results we present in the previous sections are generally robust with respect to various parameter choices.

6. Conclusion

Motivated by recent advances in the portfolio construction literature, this article proposes the use of norm-constrained portfolio optimization methodologies for efficient hedge fund clone construction. Technically the proposed method is a modification to the standard à la Sharpe 'style analysis' where the objective function is augmented with a penalty proportional to the sum of the absolute values of the replicating asset weights, i.e. the norm of the asset weights vector.

We show that the properties of the resulting portfolios are extremely attractive both relative to their benchmark hedge fund index but also relative to selected commercially available hedge fund clone products that use proprietary construction techniques. In particular the mimicking portfolios obtained through norm-constrained optimization involve a reasonable number of assets, about ten out of the initial set of twenty, and are highly correlated with the benchmark indices. Their turnover is significantly lower than that implicit in conventional techniques. Finally, they seem to provide an attractive alternative to selected commercially available hedge fund clones, especially from a risk management perspective.

References

[1]. Agarwal, V., G. Bakshi, and J. Huij, 2008, "Investment Opportunities and the Cross-Section of Hedge Fund Returns: Implications of Higher-Moment Risks for Performance," SSRN Working Paper.

[2]. Agarwal, V., W. Fung, Y. C. Loon, and N. Naik, 2006, "Risk and Return in Convertible Arbitrage: Evidence from the Convertible Bond Market," SSRN Working Paper.

[3]. Agarwal, V. and N.Y. Naik, 2004, "Risks and portfolio decisions involving hedge funds," *Review of Financial Studies* 17, 63-98.

[4]. Agarwal, V. and N. Naik, 2000, "Multi-period performance persistence analysis of hedge funds," *Journal of Financial and Quantitative Analysis* 35, 3, 327-342.

[5]. Akaike, H., 1973, "Information theory and an extension of the maximum likelihood principle," in Petrox, B.N. and F. Caski, Ed., *Proceedings of Second International Symposium on Information Theory*, Akademiai Kiado, Budapest, 67–281.

[6]. Amenc, N., S. El Bied, and L. Martellini, 2003, "Predictability in Hedge Fund returns," *Financial Analyst Journal* 59, 5, 32-46.

[7]. Amenc, N., W. Gehin, L. Martellini, and J.-M. Meyfredi, 2007, "The Myths and Limits of passive Hedge Fund Replication," Working Paper, EDHEC Risk and Asset Management Research Centre.

[8]. Amenc, N., L. L. Martellini, J. –C. Meyfredi, and V. Ziemann, 2008, "Passive Hedge Fund Replication – Beyond the Linear Case," Working Paper, EDHEC Risk and Asset Management Research Centre.

[9]. Amin, G. and H. kat, 2003, "Hedge Fund Performance (1990-2000): Do the 'Money Machines' really Add Value?" *Journal of Financial and Quantitative Analysis* 38, 251-274.

[10]. Brodie, J., I. Daubechies, C. De Mol, and D. Giannone, 2007, "Sparse and Stable Markowitz Portfolios," ECORE Discussion paper 2007/61.

[11]. Corielli, F. and M. Marcellino, 2006, "Factor based index tracking," *Journal of Banking and Finance* 30, 2215-2233.

[12]. Daubechies, I., M. Defrise, and C. De Mol, 2004, "An Iterative Thresholding Algorithm for Linear Inverse Problems With a Sparsity Constraint," *Communications on Pure and Applied Mathematics* 57, 1416–1457.

[13]. DeMiguel, V., L. Garlappi, J. Nogales, and R. Uppal, 2007, "A Generalized Approach to Portfolio Optimization: Improving Performance By Constraining Portfolio Norms," Working Paper.

[14]. Duarte, J., F. A. Longstaff, and F. Yu, 2007, "Risk and Return in Fixed Income Arbitrage: Nickels in front of a Steamroller," *Review of Financial Studies* 20, 769-811.

[15]. Dupleich, R., D. Giamouridis, S. Mesomeris, and N. Noorizadeh, 2008, "Unbundling Hedge Fund Beta," SSRN Working Paper.

[16]. Fung, W., D. A. Hsieh, T. Ramadorai, and N. Naik, 2006, "Hedge Funds: Performance, Risk and Capital Formation," SSRN Working Paper.

[17]. Fung, W., and D. A. Hsieh, 2001, "The risk in hedge fund strategies: Theory and evidence from trend followers," *Review of Financial Studies* 14, 313-341.

[18]. Fung, W., and D. A. Hsieh, 2000, "Performance characteristics of hedge funds and commodity funds: Natural vs. spurious biases," *Journal of Financial and Quantitative Analysis* 35, 291-307.

[19]. Hasanhodzic, J. and A. W. Lo, 2007, "Can hedge-fund returns be replicated?: The linear case," *Journal of Investment Management* 5, 2, 5-45.

[20]. Hoerl, A. E., and R. Kennard, 1970, "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics* 8, 27-51.

[21]. ter Horst, J.R., T.E. Nijman, and F.A. De Roon, 2004, "Evaluating style analysis," *Journal of Empirical Finance* 11, 1, 29-53.

[22]. Jaeger, L. and C. Wagner, 2005, "Factor modelling and benchmarking of Hedge Funds: Can Passive Investments in Hedge Fund strategies deliver?" *Journal of Alternative Investments*, Winter, 9-36.

[23]. Jagannathan, R., and T. Ma, 2003, "Risk Reduction in Large Portfolios: Why Imposing the Wrong Constraints Helps," *Journal of Finance* 58, 1651-1684.

[24]. Kat, H.M. and H.P. Palaro, 2005a, "Who Needs Hedge Funds? A Copula-Based Approach to Hedge Fund Return Replication," AIRC Working Paper No.27.

[25]. Kat, H.M. and H.P. Palaro, 2005b, "Hedge Fund Returns: You Can Make Them Yourself!" AIRC Working Paper No. 23.

[26]. Khandani. A. E and A. W. Lo, 2008, "What Happened To The Quants In August 2007?: Evidence from Factors and Transactions Data," SSRN Working Paper.

[27]. Liang, B., 1999, "On the performance of Hedge Funds," *Financial Analysts Journal* 55, 72-85.

[28]. Lobo, M.S., Fazel, M. and Boyd, S. (in press), "Portfolio optimization with linear and fixed transaction costs," *Annals of Operation Research*, DOI 10.1007/s10479-006-0145-1.

[29]. Mitchell, M. and T. Pulvino, 2001, "Characteristics of risk in risk arbitrage," *Journal of Finance* 56, 2135-2175.

[30]. Papageorgiou, N. A., B. Remillard, and A. Hocquard, 2008, "Replicating the Properties of Hedge Fund Returns," *Journal of Alternative Investments*, forthcoming.

[31]. Schmidt, M., G. Fung, and R. Rosales, 2007, "Fast Optimization Methods for L1-Regularization: A Comparative Study and 2 New Approaches," ECML'07.

[32]. Schneeweis, T. and H. Kazemi, 2007, "Replication of Hedge Fund Strategies," Hedge Fund Replication & Alternative Beta USA 2007.

[33]. Sharpe, W., 1992, "Asset allocation management style and performance measurement," *Journal of Portfolio Management*, Winter, 7-19

[34]. Schwarz, G., 1978, "Estimating the dimension of a model," *Annals of Statistics* 6, 461–464.

[35]. Tibshirani, R. 1996, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B.* 58, 267-288.

[36]. Vrontos, I. D. and D. Giamouridis, 2008, "Hedge Fund Return Predictability in the Presence of Model Uncertainty and Implications for Wealth Allocation," SSRN Working Paper.

[37]. Vrontos S.D., I.D. Vrontos, and D. Giamouridis, 2008, "Hedge fund pricing and model uncertainty," *Journal of Banking and Finance* 32, 741-753.

[38]. Welsch R.E. and X. Zhou, 2007, "Application of Robust Statistics to asset allocation models," *REVSTAT – Statistical Journal* 5, 1, 97–114.

Appendix

When we impose the constraint on the 2-norm, such that

$$\hat{\boldsymbol{\beta}}^{Ridge} = \arg\min_{\boldsymbol{\beta}} \left[\|\mathbf{r} - \mathbf{F}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_2^2 \right]$$

the optimization problem is still convex and the analytical solution is $\widehat{\boldsymbol{\beta}} = (\mathbf{F}^T\mathbf{F} + \lambda\mathbf{I})^{-1}\mathbf{F}^T\mathbf{r}$, where **I** is the *N*x*N* identity matrix.

Hence, introducing the 2-norm constraint allows achieving numerical stability and increased prediction accuracy. However, it does not lead to identify sparse models, since it only allows to shrink coefficient but not to have 0 values coefficients. On the contrary, imposing a constraint on the 1-norm, while sharing the advantages of imposing the 2-norm constraint, allows to select a more parsimonious and easily interpretable model since some coefficient will have 0 values. The optimization problem is an unconstrained convex optimization problem, but it is now non-differentiable when β_i=0 for any β_i . Then, we do not have a closed form solution for the global minimum but many efficient and fast techniques have recently been introduced for determining the optimal values (see Schmidt et al., 2007 for a review)

Furthermore, when we impose the constraint that $\sum_{i=1}^{N} \beta_i = 1$ and we interpret the β as the portfolio weights, we can show that such constraint allows us to control the total amount of short selling. This can be easily shown by noticing that:

$$\hat{\boldsymbol{\beta}}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \left[\|\mathbf{r} - \mathbf{F}\boldsymbol{\beta}\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1^1 \right]$$

is equivalent to:

$$\hat{\boldsymbol{\beta}}^{Lasso} = \arg\min_{\boldsymbol{\beta}} \left\|\mathbf{r} - \mathbf{F}\boldsymbol{\beta}\right\|_2^2$$

s.t. $\left\|\boldsymbol{\beta}\right\|_1^1 \leq t$

Then, if we decompose the 1-norm as:

$$\| \boldsymbol{\beta} \|_1^1 = \sum_{i=1}^{N} | \beta_i | = \sum_{i=1}^{N} (\beta_i^+ - \beta_i^-) \text{ where } \begin{cases} \beta_i^+ = \beta_i \ \ and \ \beta_i^- = 0 \ \ if \ \beta_i > 0 \\ \beta_i^+ = 0 \ and \ \beta_i^- = \beta_i \ \ if \ \beta_i < 0 \\ \beta_i^+ = 0 \ and \ \beta_i^- = 0 \ \ if \ \beta_i = 0 \end{cases}$$

We then have that:

$\| \boldsymbol{\beta} \|_1^1 \leq t$ is equivalent to $\sum_{i=1}^{N} (\beta_i^+ - \beta_i^-) \leq t$ and since $\sum_{i=1}^{N} (\beta_i^+ - \beta_i^-) = 1$, we have $-\sum_{i=1}^{N} \beta^- \leq \frac{t-1}{2}$.

Tables and figures

Table 1 Market factors

Equity	Bonds	Commodities	Currency	Real Estate/Equity
Standard & Poors 500 Total Return	Citigroup Treasury	S&P GSCI	Dollar Index	FTSE EPRA/NAREIT Global Real Estate
Russell 2000	Citigroup World Government 1-3Y	S&P GSCI Gold		
Russell 2000 Total Return	Citigroup World Government 7-10Y	S&P GSCI Crude Oil		
Dow Jones Industrial Average	Broad Investment-Grade Bond			
MSCI World Equity	Cash			
MSCI EAFE (Europe, Australasia, and Far East) Net Total Return				
MSCI Emerging Markets				
Bovespa Brazil Ibovespa				
Hang Seng				
CBOE Implied Volatility				

Table 2 Most important factors

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
EH																	
AIC			**			**			**					**			
BIC														*			
LASSO	**low**	**	**	**	**	**			**		**						
	high	**			**		**										
ED																	
AIC			**						**				**		**		
BIC			**														
LASSO	**low**	**	**	**	**			**									
	high	**			**										**		
M																	
AIC			**			**											
BIC			**														
LASSO	**low**	**	**			**				**		**					
	high	**					**						**				
RV																	
AIC								**			**						
BIC															*		
LASSO	**low**	**		**					**	**							
	high	**		**													
EM																	
AIC		**	**														**
BIC			**														
LASSO	**low**	**		**													
	high		**				**										
FC																	
AIC			**			**											
BIC			**														
LASSO	**low**	**	**	**	**	**			**								
	high	**	**		**	**	**										
FF																	
AIC			**														
BIC			**														
LASSO	**low**	**	**	**		**		**				**				**	
	high	**	**	**		**	**										

This table shows the factors that are selected by the model selection techniques. Model selection is performed through the Akaike Information criterion (AIC), the Bayesian Schwartz Information criterion (BIC), and LASSO with λ=0.5 (low) and λ=2 (high). The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF). A factor is classified as important ** when it is selected by a model in all rolling windows, i.e. 100% of the time. When there is no such factor the factor identified in most rolling windows is reported with *. The factors are: (1) the Russell 2000 Total Return Index, (2) the MSCI Emerging Markets Index, (3) the Dollar Index, (4) S&P GSCI Crude Oil Index, (5) the CBOE Implied Volatility Index, (6) the Bovespa Brazil Ibovespa Index, (7) the Hang Seng Index, (8) Cash, (9) the S&P GSCI Gold Index, (10) the FTSE EPRA/NAREIT Global Real Estate Index, (11) the S&P GSCI Index, (12) the Citigroup World Government 7-10Y Index, (13) the MSCI World Equity Index, (14) the Salomon Barney Broad Investment-Grade Bond Index, (15) the MSCI EAFE (Europe, Australasia, and Far East) Net Total Return Index, and (16) the Russell 2000 Index.

Table 3 In- and out-of-sample model fit

	Straight		**Ridge**		**AIC**		**BIC**		**LASSO**			
	In	**Out**	**In**	**Out**	**In**	**Out**	**In**	**Out**	**In**		**Out**	
									low	**high**	**low**	**high**
EH	3.43%	3.61%	3.43%	3.61%	3.46%	3.78%	3.88%	4.04%	4.02%	5.23%	3.53%	3.88%
ED	3.01%	3.40%	3.05%	3.26%	3.12%	3.18%	3.43%	3.17%	3.39%	4.26%	3.15%	3.50%
M	4.85%	5.16%	4.88%	5.24%	4.92%	5.09%	5.23%	5.33%	5.13%	6.37%	4.88%	4.75%
RV	2.46%	2.84%	2.46%	2.79%	2.53%	2.86%	4.54%	4.75%	2.70%	3.50%	2.49%	3.01%
EM	5.99%	5.85%	6.10%	5.79%	6.20%	5.42%	7.41%	5.65%	6.41%	7.41%	5.33%	5.16%
FC	2.39%	2.80%	2.42%	2.77%	2.46%	2.76%	2.84%	2.88%	2.67%	3.64%	2.63%	3.01%
FF	3.39%	3.20%	3.43%	3.23%	3.50%	3.22%	3.64%	3.47%	3.50%	4.12%	3.19%	3.33%

This table shows the in-sample (In) and out-of-sample (Out) annualised Root Mean Squared Error computed for the best model which is obtained through the alternative model selection techniques. Model selection is performed through a straight multiple regression (Straight), a ridge regression (Ridge), the Akaike Information criterion (AIC), the Bayesian Schwartz Information criterion (BIC), and LASSO with λ=0.5 (low) and λ=2 (high). The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF).

Table 4 Average model size (number of factors)

	AIC	BIC	LASSO	
			Low	high
EH	10.9	6.9	14.3	6.2
ED	7.9	5.7	13.8	9.2
M	10.1	6.6	14.5	7.4
RV	9.6	4.1	13.7	10.4
EM	9.3	1.5	14.6	3.3
FC	11.0	10.7	14.3	8.6
FF	9.3	7.3	16.0	9.7

This table shows the average model size, i.e. average number of factors, for the models identified as best by the different model selection techniques. Model selection is performed through the Akaike Information criterion (AIC), the Bayesian Schwartz Information criterion (BIC), and LASSO with λ=0.5 (low) and λ=2 (high). The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF).

Table 5 Minimum and maximum exposures to factors

	Straight		Ridge		AIC		BIC		LASSO			
	min	max	min	max	min	max	min	max	min		max	
									low	high	low	high
EH	-2.60	3.78	-0.92	3.16	-3.19	3.99	-0.98	3.43	-0.13	0.00	0.83	0.42
ED	-2.31	2.80	-1.11	2.56	-2.37	2.84	-1.12	2.76	-0.12	-0.04	1.40	0.24
M	-7.86	7.98	-2.52	2.63	-6.09	6.22	-6.25	6.36	-0.20	-0.03	0.91	0.28
RV	-3.29	3.33	-1.02	2.12	-3.88	3.92	-1.36	2.39	-0.15	-0.06	1.69	0.27
EM	-10.03	10.03	-3.59	3.61	-9.91	9.92	0.00	0.64	-0.19	0.00	0.68	0.43
FC	-2.37	2.60	-1.09	1.83	-3.40	3.65	-4.94	5.17	-0.11	-0.01	1.14	0.28
FF	-2.79	2.90	-1.39	1.85	-3.31	3.43	-3.55	3.68	-0.17	-0.07	1.36	0.16

This table shows the minimum (min) and maximum (max) exposures on the factors obtained through the alternative model selection techniques over the entire period. Model selection is performed through a straight multiple regression (Straight), a ridge regression (Ridge), the Akaike Information criterion (AIC), the Bayesian Schwartz Information criterion (BIC), and LASSO with λ=0.5 (low) and λ=2 (high). The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF).

Table 6 Mean number of assets and out of sample correlation

	No of assets			Correlation		
	Constr. Regr.	LASSO		Constr. Regr.	LASSO	
		low	high		low	high
EH	20.0	14.6**	12.5**	0.91	0.92**	0.91**
ED	20.0	15.8**	12.0**	0.84	0.85**	0.85**
M	20.0	16.7**	12.0**	0.62	0.61**	0.61**
RV	20.0	14.8**	11.8**	0.43	0.43**	0.48**
EM	20.0	15.6**	11.9**	0.91	0.91**	0.92**
FC	20.0	14.2**	10.9**	0.92	0.93**	0.92**
FF	20.0	14.9**	12.5**	0.79	0.78**	0.76**

This table shows the average number of assets in the replicating portfolio and correlation coefficients between the replicating portfolio return and the respective hedge fund index return. Replicating portfolios are constructed through constrained regression (Constr. Regr.) and LASSO with λ=0.0025 (low) and λ=0.055 (high). ** for number of assets indicates that the average number of assets is statistically different than 20, i.e. the number of assets in the portfolio obtained through constrained regression, at the 5% significance level. ** for correlation indicates correlations that are statistically different to 0 at the 5% significance level. The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF).

Table 7 Out of sample excess returns, TEV and Turnover

	Excess Returns			TEV			Turnover		
	Constr. Regr.	LASSO		Constr. Regr.	LASSO		Constr. Regr	LASSO	
		low	high		low	high		low	high
EH	-0.10%	-0.22%	-0.20%	0.39%	0.37%	0.37%	41%	14%**	8%**
ED	-0.08%	-0.31%	-0.32%	0.34%	0.34%	0.33%	21%	19%	7%**
M	0.03%	-0.12%	-0.23%	0.48%	0.48%	0.46%	36%	35%	8%**
RV	-0.08%	-0.27%	-0.30%	0.30%	0.29%	0.27%	29%	14%**	7%**
EM	-0.26%	-0.47%	-0.49%	0.57%	0.55%	0.53%	40%	33%	7%**
FC	-0.07%	-0.23%	-0.23%	0.30%	0.28%	0.28%	28%	9%**	7%**
FF	0.00%	-0.12%	-0.13%	0.32%	0.31%	0.33%	34%	12%**	12%**

This table shows the average excess returns (Excess Returns), i.e. (replicating portfolio return – hedge fund index return), tracking error volatility (TEV), i.e. the standard deviation of the excess return, and average turnover of the replicating portfolio. Replicating portfolios are constructed through constrained regression (Constr. Regr.) and LASSO with λ=0.0025 (low) and λ=0.055 (high). ** for excess returns indicates that excess returns from portfolios obtained from constrained regression and lasso low/lasso high come from distributions with statistically unequal means at the 5% significance level. ** for turnover indicates that the turnover of portfolios obtained from constrained regression and lasso low/lasso high come from distributions with statistically unequal means at the 5% significance level. The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF).

Table 8 Extreme allocations

	Constr. Regr.			LASSO					
				low			high		
	no factors	average	max	no factors	average	max	no factors	average	max
EH	10	61.58%	100.00%	2	5.45%	9.90%	0	-	-
ED	10	77.03%	100.00%	2	39.60%	78.22%	2	12.87%	14.85%
M	11	61.66%	100.00%	9	20.57%	37.62%	0	-	-
RV	8	60.15%	100.00%	4	18.07%	48.51%	2	26.24%	29.70%
EM	11	71.56%	100.00%	7	15.28%	30.69%	1	1.98%	1.98%
FC	10	55.64%	100.00%	0	-	-	0	-	-
FF	9	54.46%	100.00%	0	-	-	0	-	-

This table shows the number of assets in the replicating portfolios with weights greater than 50% either long or short. In column headed ‘no factors’ we present the number of factors which had a high loading, at least once, during our sample period. In column headed ‘average’ we present what proportion of the test period on average has the loadings on these factors been high. The column head ‘max’ presents the max proportion of the test period that a factor has exceeded the threshold. Replicating portfolios are constructed through constrained regression (Constr. Regr.) and LASSO with λ=0.0025 (low) and λ=0.055 (high). The dataset covers the period February 1990 to July 2008 and considers the four HFR primary strategy indices namely Equity Hedge (EH), Even Driven (ED), Macro (M), and Relative Value (RV), an index which comprises managers with regionally focus the Emerging Markets (EM), an aggregate Hedge Fund market index, the Fund Weighted Composite Index (FC), and a proxy for the aggregate Fund of Funds market, the Fund of Funds Composite Index (FF).

Table 9 Performance relative to selected commercial products (Monthly returns)

	FC	FC_{CR}	FC_{LASSO} low	FC_{LASSO} high	FC_{ML}	FF	FF_{CR}	FF_{LASSO} low	FF_{LASSO} high	FF_{CITI}	FF_{DB}
Mean return (%)	0.86	0.86	0.72	0.68	0.61	0.64	0.71	0.60	0.57	0.83	0.76
Std (%)	1.42	1.54	1.51	1.51	1.44	1.29	1.16	1.11	1.09	1.77	2.08
Mean return/ Std	0.61	0.56	0.47	0.45	0.43	0.50	0.61	0.54	0.52	0.47	0.36
Correlation with benchmark	-	0.91	0.91	0.90	0.91	-	0.82	0.81	0.80	0.78	0.79
VaR95 (%)	1.72	1.79	1.88	2.03	1.57	1.97	1.40	1.47	1.38	2.00	2.57
VaR99 (%)	2.62	3.15	3.17	3.04	5.20	2.87	2.63	2.43	2.25	3.38	8.00
CVaR95 (%)	2.38	2.58	2.73	2.65	3.51	2.59	2.13	2.11	2.00	2.86	5.72
CVaR99 (%)	2.69	3.30	3.27	3.15	5.64	2.90	2.75	2.49	2.30	3.43	8.33

This table shows performance statistics for replicating portfolios and hedge fund indices, i.e. the Fund Weighted Composite Index (FC), and the Fund of Funds Composite Index (FF). The dataset covers the period February 2003 to July 2008. The statistics are the: average return (Mean return), standard deviation of returns (Std), the ration of average returns over their standard deviation (Mean return/Std), absolute Value-at-Risk at the 5% level (VaR95), absolute Value-at-Risk at the 1% level (VaR99), absolute Conditional Value-at-Risk at the 5% level (CVaR95), and absolute Conditional Value-at-Risk at the 1% level (CVaR99). Replicating portfolios are constructed through constrained regression (CR) and LASSO with λ=0.0025 (low) and λ=0.055 (high). Selected currently available products include Citi's HARP Index (FF_{CITI}), Deutsche Bank's Absolute Return Beta Index (FF_{DB}), and Meryll Lynch's Factor Index (FC_{ML}).