Insight – Distributional Hedge Fund Replication via State Contingent Stochastic Dominance

Clemens H. Glaffig

(appeared in: "Hedge Fund Replication", ed. G. Gregoriou, M. Kooli, Palgrave Macmillan, 2012)

Abstract

We propose a new hybrid hedge fund replication technique, which combines aspects of portfolio construction from factor based replication with an innovative version of distributional replication. It uses a parameterized replicating strategy for which we match a version of state contingent integral stochastic dominance. The dominance will be over a set of distributions reflecting preferred, state contingent distributional performance characteristics, granting insight into return features to arbitrary fine detail. It emphasizes the replication of desired aspects rather than the replication of the performance path. A further application of this approach is to replicate certain features of a target fund and at the same time dominate on less desired aspect. Before the new approach to replication is presented, a brief recollection of the evolution, the various different approaches and some of the pitfalls of hedge fund replication are highlighted.

1. Introduction

The topic of hedge fund replication is well established, both in the academic world with a growing list of literature as well as in applications with an active market for replication- and alternative beta products. Its theoretical concepts translated into market applications early on, generating wide interest among investors, even though the resulting products were not always performing to expectations. In section 2, we will comment on the evolution of hedge fund replication, including a view as to the role replication techniques can play in applications going forward: Not constructing simple clones but improving them on specific performance characteristics. In section 3 we give brief descriptions of previous replication techniques. In section 4 we highlight some of the pitfalls of common replication techniques. Section 5 will introduce a hybrid replication technique, combining the intuitive portfolio construction of factor based replication with the less ambitious goal to replicate certain preferred and state contingent characteristics of the return distribution rather than replicating the performance path. This approach is specifically geared to accommodate the role of replicating the good- and improving on the bad performance aspects of a given hedge fund target. In section 6 we give an empirical application of the proposed replication technique and compare it to factor based regression, the most common replication technique thus far. Section 7 concludes.

2. The Evolution of Hedge Fund replication

The end of the bull market in equities at the turn of the century presented investors with unaccustomed volatility and excessive losses. It resulted in rethinking some aspects of the classical approach to asset management of benchmark driven investments, diversifying into benchmark free absolute return strategies.

Hedge funds, thus far a rather mystical and secretive class of investment strategies granting little insight and reporting infrequently, rose sharply in popularity with main stream investors. Opening the segment to more traditional investors, the need to understand, or at least shed some light on the inner workings of hedge funds arose.

The first efforts in hedge fund replication developed in that context, motivated by the desire to detect and understand the main risk factors driving hedge fund performance, by regressing historic hedge fund returns against the return of a set of style factors.

While the initial attempt was to reproduce the historic performance path of a given target fund, it was soon realized, that these techniques could also be used to replicate hedge fund returns on a forward looking basis. The vision was to construct a recipe for allocating funds within a small universe of liquidly tradable instruments exhibiting identical performance behavior than the target fund but being fully transparent, highly liquid and with substantially lower fees than what the target fund would charge.

The original hype and hope quickly disappeared, as it became apparent that the ability to successfully track a historic performance path does not necessarily translate into producing enough insight into the true nature of the targeted strategy to construct a clone. The dynamic features, specific trader’s talents, the granularity achieved by successful multi strategy funds or other idiosyncrasies are too often just to dominating a factor in hedge fund performance characteristics and not really adept to simple, semi static modeling. Consequently, the results were mixed at best.

To address these deficiencies hedge fund replication went on to concentrate on tracking indices. The hope was that for indices, idiosyncrasies average out, more systematic and easy to clone features of performance characteristics would dominate. Here, the results were more promising, relying however very much on combining the right instruments and trading rules within the replicator to capture the dynamic and non linear aspects that are still prevalent in style indices; see e.g. Gupta, Szado and Spurgin (2008), Amenc et al. (2008).

To relax on the ambitious objective of path replication, new replication techniques like rule based trading- or distributional replication targeted to capture and replicate general, essential features of specific styles, referring to them as alternative beta strategies with no or little ambition of cloning the performance path anymore; see e.g., Kat and Palaro (2005) for distributional replication. For a recent comparative survey of replication products in the market, see e.g. Tuchschmid, Wallerstein and Zaker (2010).

Competing with replication products are fund of hedge funds and more recently portfolios of real hedge funds in the form of ETFs, promoted by larger providers of managed account platforms with the advantage of cost efficient access. Exhibit 1 illustrates the daily performance graphs of the HFR General Hedge Fund Index against three replication examples: The GS Absolute Return Tracker, the JP Alternative Beta Reference and the ML Factor Model Index.

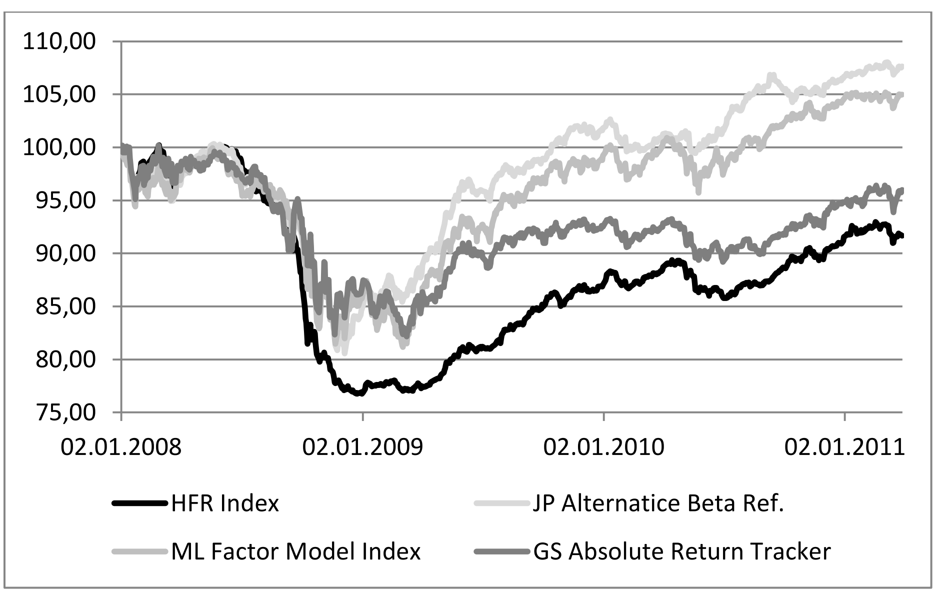

Exhibit 1: Performance Graph for the daily performance from January 2, 2008 to March 31, 2011.

The role of hedge fund replication

Replicating the essential performance features of a style is nowadays the objective followed by most applications of hedge fund replications in the market. As products that capture and reflect the essential performance characteristics of general styles, they provide a liquid, transparent and cheap basis to construct a portfolio of hedge funds within a core-satellite approach. They simplify and speed up the allocation to different style sectors and even provide means to shorten essential aspects of specific style performance attributes, serving as an efficient risk management tool for portfolios of hedge funds.

The scope could be extended to provide a basis for what investors really want and expect from style related alternative beta strategies: A product that reflects the sum of all good attributes of a hedge fund style, with some of the undesired elements (tail correlation, etc.) removed. Not style clones but products that dominate style indices based on bespoke criteria, while still exhibiting major performance characteristics of the chosen style. The reflected bespoke preferences could go well beyond correlation to existing portfolios or preferences with respect to moments of the return distribution.

3. Replication techniques

Most replication techniques belong to one of two main classes: Factor based- and distributional replication. In addition there are a number of hybrid- as well as individual approaches.

Factor based replication

Factor based replication, the most common replication technique, is based on Sharpe’s concept of style factors to capture the essentials of an active management of classical strategies. Fung and

Hsieh (Fung and Hsieh, 1997) extended this approach to hedge funds and hedge fund styles. Factor based replication is in general a parameterized, rule based strategy, in which the parameters are adjusted to best replicate or track a given performance path over a specific history. It not only tries to match returns, but also the sequential order in which returns are realized.

In the most common case, factor replication is a multilinear regression, the parameters just the factor loadings or the betas with respect to standard investable market risk factors; the produced rules are consequently just the static allocations according to these betas. More elaborate models include option based strategies and alternative indices reflecting certain rule based trading strategies as factors; see e.g. Fung and Hsieh (1997) &(2002), Jaeger and Wagner (2005) , Hasanhodzic and Lo (2007), Spurgin (1999), Gehin, Martellini, MeyFredi (2007) and references therein.

The advantage of factor based replication lies in its simplicity. The art and value added of this approach rests with the specification of the factors: The quality of the replication depends heavily on capturing the dynamic aspects including carry aspects and tail event behavior, which can be improved by including alternative indices reflecting non standard trading strategies (like, e.g., the BMX index) or individually defined rule sets.

Distributional replication

Distributional hedge fund replication is the approach advocated by Kat and Palaro. It is based on the dynamic replication developed in the framework of contingent claim valuation. For a given portfolio, it results in matching the joint distribution of the replicator and the portfolio to that of the target and the portfolio on a percentile by percentile basis; see e.g. Kat and Palaro (2005) & (2007), Gehin, Martellini, and MeyFredi (2007).

In contrast to factor based replication, its goal is not to approximate the performance path, but distributional aspects, specifically frequency counts of returns, regardless of the sequential order these returns are realized. It is thus less ambitious than factor based replication. Distributional replication is potentially less prone to over fitting, while it may result in completely different realizations of performance paths. The implementation however is clearly more complicated than factor based replication.

Hybrid theory and other alternatives

Further approaches combine factor based models with aspects of distributional replication, see e.g. Kazemi (2007) and section 5 below. Other examples use, e.g., tracking techniques well established in various nonfinancial applications based on MCMC methods that may be suited to deal with non stationarities; see Roncalli and Weisang (2008).

4. Some pitfalls in replication

Using hedge fund replication techniques to produce cheap and transparent clones of hedge funds has thus far only been moderately successful. Plenty of explanations for this lack of success have been given. The most obvious deficiency of common replication techniques is the lack of dynamics and the difficulty to model idiosyncrasies ("trader's talent"). More generally, any non-stationary trading behavior will be difficult to model. While these problems weaken somewhat by following less ambitious replication objectives or averaging out idiosyncrasies by targeting indices instead of single funds, a number of challenges remain, that can, even if the objective is only to gain insight and understanding of past performance, result in misleading conclusions. We will highlight and repeat some-, but certainly not all of the pitfalls of hedge fund replication.

Over fitting of factor based replication: As with all parametric approaches, adding ever more factors to a replicator will clearly improve in-sample tracking performance but is very prone to over fitting. Capturing the dynamics of the target necessitates the inclusion of additional nonstandard factors, deceiving into overloading on different factors that only marginally contribute to explanation. Prudent pruning is required. Closely related to this is:

Factor misspecification: Potential factor misspecifications are manifold. The desire to capture the detail, the dynamics and nonlinear performance behavior of hedge funds often leads to the inclusion of a large number of highly inter-correlated factors. If those factors don't play a role in the target fund's strategy, the individual factor loadings will often not optimize down to zero but produce combinations of long and short positions that have no relevance to the target, resulting not only in over fitting, but also in a misconception of the target's risk and strategy. As another specific example, successful market timing within the observed data frequency does create alpha. Sometimes, for very pronounced timing talents, this may look like being long a lookback option. Often though, this alpha looks very much like option selling or some other positive carry strategy on the larger observation scale. Typically, factor based replication will model this with an option selling factor, exhibiting an asymmetric, negatively skewed performance behavior, which the target may not have.

Misspecification of the distribution: Distributional replication depends very heavily on getting the distribution right. Real life distributions are typically approximated by well behaved and easy to handle distribution classes. Given sparse data like monthly returns, these approximations are very crude and poorly separate different strategies. Fitting multivariate copulas to model multivariate distributions cannot be properly done based on the available data history of most funds or indices.

Incomplete data: Parameters of replication models are optimized throughout some historical period. If those are incomplete to the extent, that e.g. real tail events are not included, behavior in extreme market phases will not be reflected. Consequently, risk management and stop loss behavior are not adequately modeled. In general, the data used may be insufficient to properly separate different strategies leading to similar behavior in the given period, but potentially big performance differences in market phases not covered by the data used to optimize parameters. This leads to misspecifications specifically for rule based trading replications.

Data frequency: An increasing number of funds trade at high frequency intervals. Modern statistical arbitrage has evolved to a good degree into high frequency trading. Others like short term CTA's close out or minimize their risk at the end of each trading day. Working with monthly return data to understand the aspects of such trading strategies is meaningless. Even if more funds, specifically under the format of UCIT III in Europe report daily performance numbers, high frequency strategies or indices with a large component thereof will be very difficult to replicate without very high misspecification risk. Data frequency will also influence the factor choice: If trading frequency differs systematically from observation frequency - e.g. daily trading, monthly observations – the trading range and relative closing level of factors will become more important than absolute closing levels.

5. A new hybrid approach – dominance replication

In the following we will highlight a hybrid approach of factor based- and distributional replication.

The distributional replication as developed by Kat and Palaro (see Kat and Palaro, 2005) in a series of papers is in general a two step process, based on the replication of contingent payoffs as initially developed by Merton in the context of Black-Scholes option pricing theory: For a given fund, H, a so called reserve asset, R, which drives the replication and an initial portfolio or benchmark B, a payoff function g is replicated, s.th. for r_B, r_R and r_H denoting the respective period returns for B, R and H

$$P(g(r_R, r_B) < u_1, r_B < u_2) = P(r_H < u_1, r_B < u_2), \forall u_1, u_2 \in [0,1] \qquad (1)$$

(see also Gehin, Martellini and MeyFredi, 2007).

The replication of g follows the general theory of replicating contingent claims.

If we identify the payoff function g with its replication X, i.e. X = g(R,B) and $r_X = g(r_R, r_B)$, (1) is equivalent to

$$P(g(r_R, r_B) < u_1 \mid r_B < u_2) = P(r_H < u_1 \mid r_B < u_2), \forall u_1, u_2 \Leftrightarrow F_{X|B}(u) = F_{H|B}(u), \forall u \qquad (2)$$

where $F_{X|B}, F_{H|B}$ denote the conditional distribution function of X respectively H given B.

(2) states that the conditional distribution functions of the replication and the fund are matched. Another way to express this is to say, they have identical state contingent first order stochastic dominance properties.

We will take this view to propose an alternative way for replication, which we will refer to as dominance replication. Our motivation for a new approach is threefold:

- We would like to replicate preference based state contingent distributional performance characteristics of a specified fund or index to arbitrary fine detail.
- We would like the intuitive construction of a state dependent replicating strategy as given by factor replication, i.e. the allocation into various risk factors in a state contingent way.
- We would like to limit the potential of over fitting that is often the price to be paid by adding too many parameters and incorporating dynamic strategies.

We will express state contingent performance characteristics by considering the dominance of the state conditional distribution of returns over target distributions, which reflect and represent these characteristics in a state contingent way: Let B denote some benchmark, which could be an index, a given portfolio or a general state indicator of the market. A realization of B will be denoted a state; state contingency will mean conditioning on B = b for some specific state b. Let $\Omega = \Omega(B)$ be a family of one dimensional conditional target distributions, conditioned on B, reflecting the specific conditional performance preferences with $|\Omega| = N$. Each $T \in \Omega$ will describe a state contingent distributional target characteristic, represented by its respective conditional target distribution $F_{T|B=b}$. A fund X will be measured against these target characteristics by evaluating a version of integral stochastic dominance of the conditional distribution function $F_{X|B=b}$ over $F_{T|B=b}$. If two funds X and H have the same dominance values for a given set of target characteristics, they will be viewed to have the same state contingent distributional performance characteristics (relative to the chosen target set).

Example: Let T describe a skew target relative to the benchmark B. This could be represented by a distribution that, for each realization b of B, is scattered around b with conditional expectation exceeding b by some $\varepsilon > 0$, i.e., $E(T \mid B = b) = b + \varepsilon$. Two funds X and H with the same dominance value over T will have the same conditional skew characteristic.

Definition

For a singlet T, we call $\Delta_{T,B}(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(F_{T|B}(x \mid b) - F_{X|B}(x \mid b))dF_{T|B}(x \mid b)dF_B(b)$

the state contingent dominance of X over the target T, contingent on the benchmark B.

We call

$$\Delta_{\Omega,B}(X) = \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (F_{T_n|B}(x|b) - F_{X|B}(x|b)) dF_{T_n|B}(x|b) dF_B(b), n = 1,2,..., N \right\} \in \mathrm{R}^N$$

$$= \left\{ \Delta_{T_n,B}(X), T_n \in \Omega, n = 1,2,...N \right\}$$

the state contingent dominance vector of X over the performance target class Ω *contingent on the benchmark B.*

The vector $\Delta_{\Omega,B}(X)$ reflects to what extent X has better distributional performance characteristics than the target characteristics we have imposed by defining Ω.

With this, we will call X a dominance replication of H if $\Delta_{\Omega,B}(X) = \Delta_{\Omega,B}(H)$. Given a set Ξ of available strategies, we call X_0 the best dominance replication under Ω if:

$$X_0 = \arg\min_{X \in \Xi} \left\{ \| \Delta_{\Omega,B}(X) - \Delta_{\Omega,B}(H) \| \right\}$$

Matching contingent dominance vectors seems to be less ambitious than matching conditional distribution functions, as this is equivalent to matching first order stochastic dominance properties. However, as we control the set Ω, distributional performance characteristics of arbitrary fine detail can be reflected. With this flexibility, distributional replication as defined by (1) is a special case of dominance replication:

Let $\Omega := \left\{ T, T_{|B=b_0} = x_0, b_0, x_0 \in R \right\}$ i.e. Ω comprises the set of all "double delta" distributions $dF_{T|B=b} = \delta_{x_0}(x)\delta_{b_0}(b)$. Matching the dominance set for X and H is equivalent to $F_{X|B=b_0}(x_0) = F_{H|B=b_0}(x_0), \forall b_0, x_0$, which is (2).

The replicator

We will consider replicators from a set of parameterized strategies. Like in factor based replication, we choose a set of ABSFs R_i. The replicator set is then defined by

$$\mathrm{R} := \left\{X, X = \sum \beta_i R_i, i = 1,2...,M\right\}$$

where the β_i satisfy some normalization condition and are allowed to be state and path dependent to allow for, e.g., stop losses and CPPI-like features.

The dominance-replication strategy will then be the strategy in the replicator set satisfying:

$$X_0 = \underset{X \in \mathrm{R}}{\arg\min}\left\{\| \Delta_{\Omega,B}(X) - \Delta_{\Omega,B}(H) \|\right\} = \underset{X \in \mathrm{R}}{\arg\min}\left\{\sum_k |\Delta_{T_k,B}(X) - \Delta_{T_k,B}(H)|, k = 1,2...|\Omega|\right\}$$

While dominance replication is multilinear in the ABSFs, it is different to factor based multilinear regression: The optimization of the parameters is not done by minimizing the squared distance of the respective returns, but by minimizing the distance of conditional dominance vectors reflecting bespoke characteristics of the return distribution.

Motivation and value added for dominance replication

One of the goals to be achieved by a novice technique in hedge fund replication is to improve on the deficiencies of previous approaches. Factor based replication tries to match the performance path and hence the sequential returns, i.e. size and order of the realized returns. This stringent objective easily leads to over fitting and factor misspecification. Dominance replication tries to match some chosen performance characteristics in a weak distributional sense instead, reducing the risk of over fitting. Given that it matches integrated conditional distributions, it is less sensitive to misspecification of the distribution as in the case of distributional replication, which matches cumulative distribution functions point by point. Dominance replication can retain some aspects of path dependence by matching dominance over a whole set of state conditional distribution targets, each single target adding an anchor with respect to sequential ordering. In addition, dominance replication extends easily to an approach for constructing superior strategies in the sense that they dominate given funds, styles or indices on a bespoke performance characteristics basis. As a drawback, implementing dominance replication is clearly more elaborate than factor based replication.

6. Empirical applications

We apply dominance replication to track the HFR Equity Hedge Index and compare the resulting in sample and out of sample results on a daily basis to the standard factor based replication, using the same ABSFs for both cases. The HFR Equity Hedge Index provides delayed daily data back to 2003. We optimise parameters for both approaches over a three year period, i.e. over roughly 750 data points. We will then let the resulting replicator run with the optimized parameters for the next month, i.e. roughly 20 trading days, at which point we will repeat the optimization procedure to readjust the parameters. The initial starting date for the out of sample period is January 1, 2010. The period for determining the parameters ranges from January 1, 2006 to December 31, 2009. We will subsequently move both the training period as well as the out of sample starting point forward by one month each for the following 12 months, s.th. the procedure produces 12 months of daily out of sample returns.

The ASBFs we use are: S&P 500, a CPPI strategy on the S&P 500, VIX Index, BMX index of covered call writing, MSCI EM & EAFE, Russell2000, Russell 1000 Growth, Russell 1000 Value. We use constant factor loadings.

For the dominance replication the benchmark we use is a market state indicator, based on the S&P 500, distinguishing 7 states: {very negative, negative, slightly negative, neutral, slightly positive, positive, very positive}, defined via quantiles of the 5-day S&P 500 returns. To estimate the conditional distribution of replicating strategies, we use kernel based approximation of the empirical conditional distribution – there are enough data points for all individual realizations of the state indicator to justify this approach. The state-conditional target distributions against which we match the respective dominance of replicator and HFR index are chosen as:

$T_0 |_{B=b} (x) \sim \mathrm{N}(0,\sigma)\ \forall \mathrm{b}$ for the degree of dominance over state independent pure random scattering around zero return, where $\mathrm{N}(0,\sigma)$ denotes the Gaussian distribution with zero mean an variance σ.

$$T_k^1 |_{B=b} (x) \sim \mathrm{N}(0,\sigma)\, 1_{(b=\{k\})}, k = 1,.2,...7$$

for the degree of dominance over state contingent, pure random scattering around zero return, separately for each market state.

$$T_k^2 |_{B=b} (x) \sim 0.5[\{\text{"negative tail event"}\} + \{\text{"positive mean event"}\}]1_{(b=\{k\})}, k = 1,.2,...7$$

for the degree of dominance over state contingent tail skews, separately for each market state.

Results:

The results for the out of sample replications are summarized in tables 1 and 2. The out of sample r-squared of the standard factor based regression with daily return data is only 12%, while the out of sample r-squared of the dominance replication is 65%. Improvements, specifically for the regression can be obtained by de-correlating the factor set. However, as can be seen from exhibit 2, the tracking result for the regression is optically not nearly as bad as the r-squared would suggest: A lot of investors can live with such a replication. Nevertheless, if the objective is to track the index in some sense, the tables and the exhibit indicate that dominance replication has captured the essentials and gained insight into the inner workings of the index better than regression replication.

	Correlation		R-Squared	
	daily	monthly	daily	monthly

Dominance	1%	61%	65%	66%
Regression	-1,4%	84%	12%	23%

Table 1: Goodness of fit measures, using daily and monthly out of sample return data for the period January 1, 2010 to December 31, 2010.

	Mean	Stand.Dev.	Skew	Kurtosis
Dominance	0.026%	0.588%	7.98	3.242
Regression	0.017%	0.257%	-42.71	1.956
HFR Index	0.024%	0.401%	-13.02	1.007

Table 2: Moments for the out of sample daily return series for the period January 1, 2010 to December 31, 2010.

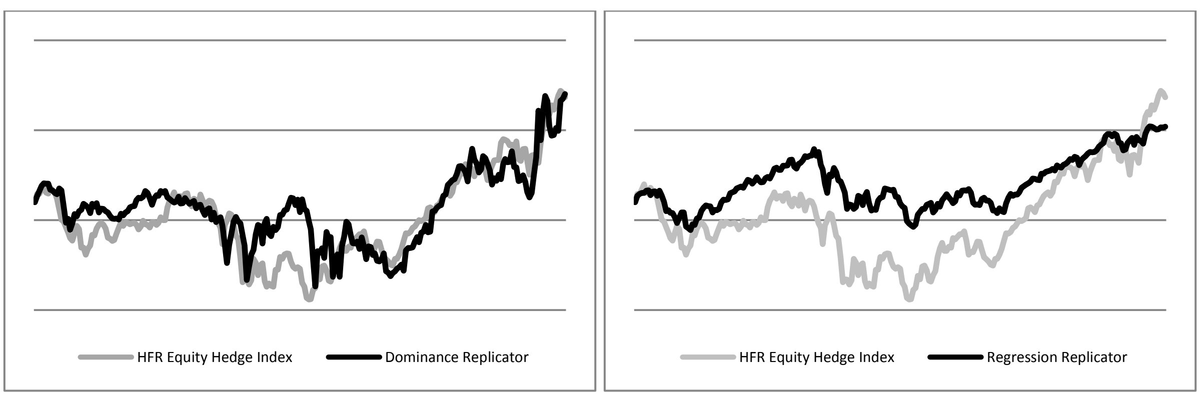

Exhibit 2: Out of sample performance graphs for the period January 1, 2010 to December 31, 2010.

7. Conclusions

Dominance replication is a parameterized replication strategy in which parameters are optimized not by trying to replicate the performance path but by replicating preferred state contingent performance characteristics. It concentrates on aspects of the return distribution that are important for an individual application and can be tuned to arbitrary fine detail, which makes it more general than classical distributional replication. Dominance replication can easily be extended to go one step beyond pure replication: Replicate desired-, improve and dominate undesired characteristics.

References

Amenc, V; Gehin, W.; Martellini, L.; MeyFredi, J. (2008). "Passive Hedge Fund Replication – A Critical Assessment of Existing Techniques." *The Journal of Alternative Investments. Fall 2008.*

Fung, W.; Hsieh, D. (1997). "Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds." *The Review of Financial Studies, Vol 10, No. 2.*

Fung, W.; Hsieh, D. (2002). "The Risk in Fixed Income Hedge Fund Styles." *Journal of Fixed Income, Vol. 12, No. 2.*

Gehin, W.; Martellini, L.; MeyFredi, J. (2007). "The Myths and Limits of Passive Hedge Fund Replication." Edhec Position Paper, Edhec Business School.

Gupta, R.; Szado, E.; Spurgin, W. (2008). "Performance Characteristics of Hedge Fund Replication Programs." Working Paper, CIDSM.

Hasanhoddzic, J; Lo, A. (2007). “Can Hedge Fund Returns be Replicated?: The Linear Case”. *Journal of Investment Management, Vol. 5, No. 2.*

Jaeger, L.; Wagner, C. (2005). „Factor Modeling and Benchmarking of Hedge Funds: Can Passive Investment Strategies Deliver?“ *The Journal of Alternative Investments. Winter (2005).*

Kat, H.; Palaro, H. (2005). „Who Needs Hedge Funds? A Copula-Based Approach to Hedge Fund Return Replication“. Working Paper #027, Alternative Investment Research Centre Working Paper Series, Cass Business School, City University, London.

Kat, H.; Palaro, H. (2007). „Alternative Routes to Hedge Fund Return Replication“. Working Paper #037, Alternative Investment Research Centre Working Paper Series, Cass Business School, City University, London.

Kazemi, H. (2007). “A Note on the Replication of Properties of Hedge Fund Returns.” Working Paper, Isenberg School of Management, University of Massachusetts, Amherst.

Roncalli, T.; Weisang, W. (2008). “Tracking Problems, Hedge Fund Replication and Alternative Beta.” Working Paper, Available at: http://ssrn.com/abstract=1325190.

Spurgin, R. (1999). „A Benchmark for Commodity Trading Advisor Performance.“ *The Journal of Alternative Investments. Summer (1999).*

Tuchschmid, N.; Wallerstein, E.; Zaker, S. (2010). “Hedge Fund Clones are Still to Count on.” Working Paper, Available at: http://ssrn.com/abstract=1729109.