This article was downloaded by: [155.246.103.35] On: 05 April 2017, At: 23:17
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Evaluating Hedge Funds with Pooled Benchmarks

Michael S. O'Doherty, N. E. Savin, Ashish Tiwari

To cite this article:
Michael S. O'Doherty, N. E. Savin, Ashish Tiwari (2016) Evaluating Hedge Funds with Pooled Benchmarks. Management Science 62(1):69-89. http://dx.doi.org/10.1287/mnsc.2014.2056

Full terms and conditions of use: http://pubsonline.informs.org/page/terms-and-conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2016, INFORMS

Please scroll down for article—it is on subsequent pages

INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

MANAGEMENT SCIENCE
Vol. 62, No. 1, January 2016, pp. 69–89
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.2056
© 2016 INFORMS

Evaluating Hedge Funds with Pooled Benchmarks

Michael S. O'Doherty
Department of Finance, Trulaske College of Business, University of Missouri, Columbia, Missouri 65211,
odohertym@missouri.edu

N. E. Savin
Department of Economics, Tippie College of Business, University of Iowa, Iowa City, Iowa 52242,
gene-savin@uiowa.edu

Ashish Tiwari
Department of Finance, Tippie College of Business, University of Iowa, Iowa City, Iowa 52242,
ashish-tiwari@uiowa.edu

The evaluation of hedge fund performance is challenging given the flexible nature of hedge funds' strategies and their lack of operational transparency. As a result, inference about skill is inevitably contaminated by the error in the benchmark model. To address this concern, we propose a model pooling approach to develop a fund-specific benchmark obtained by pooling a set of diverse attribution models. The weights assigned to the individual models in the pool are based on the log score criterion, an information-theoretic measure of the conditional performance of a model. We illustrate the advantages of a pooled benchmark over alternative approaches, including the Fung and Hsieh [Fung W, Hsieh DA (2004) Hedge fund benchmarks: A risk-based approach. *Financial Analysts J.* 60:65–80] model, stepwise regression methods, and style-adjusted methods in the contexts of a real-time investment strategy, hedge fund replication, and fund failure prediction.

Keywords: hedge funds; performance evaluation; model pooling; model combination; hedge fund replication; log score
History: Received December 13, 2013; accepted August 19, 2014, by Wei Jiang, finance. Published online in *Articles in Advance* February 13, 2015.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

1. Introduction

The evaluation of hedge fund performance continues to be a challenge for investors and researchers alike. By their very nature hedge funds' strategies are quite flexible in terms of their asset class exposure, leverage, and choice of markets in which to operate. Further, the lack of operational transparency and performance reporting requirements makes it difficult to pin down the investing style followed by a particular hedge fund. Consequently, there is no clear consensus in the literature about the appropriate model to be used for evaluating hedge funds. It is not atypical for studies in this area to employ models with anywhere from seven to a dozen or more potential factors. More importantly, the use of a particular model to the exclusion of other candidate models may be suboptimal and potentially misleading when all models represent approximations to reality and none is literally true. Instead of relying on model selection mechanisms to identify a single model to evaluate individual hedge fund performance, we propose a model pooling approach.[1] Specifically, the paper proposes a fund-specific benchmark obtained by pooling a set of diverse performance attribution models. The objective of this research is to demonstrate the advantages of model pooling in both statistical and economic terms when evaluating fund performance.

The recent global financial crisis of 2008 and the well-publicized failures of several prominent funds have led to a renewed focus on the issue of model error. In this context, there is debate in the literature surrounding how to assess funds that display low R^2 values with respect to conventional factor models. Titman and Tiu (2011) and Sun et al. (2012), for example, suggest that skilled fund managers tend to hedge systematic risk or follow distinctive strategies and therefore exhibit low R^2 values in factor model regressions. In contrast, Bollen (2013) shows that although low (effectively zero) R^2 funds have high alphas, they also exhibit a high probability of subsequent failure. Such funds also appear to have significant exposure to one or more systematic factors omitted from traditional attribution models. This debate underscores the need for more robust evaluation methods for hedge fund performance.

Typically, performance evaluation of hedge funds relies on one of three approaches. The first approach

[1] The conventional practice is to either adopt a common benchmark model that incorporates asset-class-based factors reflecting the style of the hedge funds being evaluated or to construct fund-specific benchmarks using a model selection technique like stepwise regression.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

involves the use of a single model, for example, the Fung and Hsieh (2004) seven-factor model, as an attribution model for all hedge funds. Given the flexibility of hedge funds' strategies, identification of the relevant set of factors that are universally applicable is quite challenging. By its very nature, such an approach is almost guaranteed to result in an attribution model that is misspecified for many funds. An alternative approach is to rely on a model selection technique such as stepwise regression to identify a subset of fund-specific factors from an initial set of prespecified factors. This approach is also problematic in view of data limitations. Given the relatively short return histories for many hedge funds, one is forced to consider only a small subset of factors for each fund. Moreover, a major concern with this approach is that factor selection is based on in-sample fit, which in turn leads to the problem of overfitting and the choice of factors with poor explanatory power in subsequent periods. A third alternative is to employ a style-specific benchmark by comparing a fund's return with all hedge funds in the same style category. This simple approach has the advantage that it makes a minimal demand on the length of return series required for evaluation but may not adequately account for within-style differences in fund risk exposures.

In contrast to the conventional practice, this paper employs a pooled benchmark that is customized to each individual hedge fund's style. The pooled benchmark model is formed by optimally combining five linear, factor-based attribution models that are designed to reflect diverse hedge fund strategies. The five strategy-coherent factor subsets include (a) five domestic equity factors, (b) three fixed income and commodity factors, (c) a set of three global factors, (d) the five Fung and Hsieh (2001) trend following factors, and (e) the four Agarwal and Naik (2004) option-based factors. In all, the five models include 20 distinct factors. The weights assigned to the individual factor subsets in the optimal pool are based on the log score criterion following the approach of Geweke and Amisano (2011) and O'Doherty et al. (2012). Intuitively, the log score is a measure of the conditional performance of a factor model in terms of its ability to track the monthly return for a given hedge fund.

Conceptually, the use of a pooled benchmark offers a number of advantages over the alternative approaches mentioned above. First, the pooling approach implicitly recognizes that all factor models are subject to error. The optimal combination of the distinct sets of factors helps diversify the specification error inherent in any one set of factors. Second, by pooling a number of strategy-coherent factor subsets, the optimal pooled benchmark potentially incorporates information from all 20 factors. Given the short histories of hedge funds, this is not feasible under the conventional factor selection approaches. Third, in contrast to factor selection methods like stepwise regression, the optimal pooling procedure guards against overfitting by downweighting factors/models that do not perform well out of sample. Finally, by assigning fund-specific weights to each of the strategy-coherent factor subsets, the pooled benchmark yields an economically meaningful decomposition of a fund's investment strategy.

Our primary analysis is based on the sample of hedge funds in the Lipper Tass database during the period from January 1994 to December 2011. We first show that pooled benchmarks perform significantly better than individual models in capturing the investment styles of hedge funds being evaluated. For example, based on the log score criterion, a measure of the conditional performance of a model, the optimal pooled benchmark outperforms the Fung and Hsieh (2004) seven-factor model for 88% of the sample funds. The pooled benchmark also outperforms, approximately 96% of the time, an alternative strategy of using a customized benchmark that is based on a stepwise regression approach. We further assess model performance on the basis of out-of-sample mean square tracking errors and find similar results.

We next characterize the gains to employing a pooled benchmark by considering a real-time strategy of annually sorting hedge funds into decile portfolios. Funds are sorted on the basis of their historical performance over the prior 24 months, as indicated by the t-statistics corresponding to each of the following performance measures: (a) pooled alphas, i.e., alphas computed with respect to the pooled benchmark, (b) Fung and Hsieh (2004) model alphas, (c) stepwise regression model alphas, and (d) style-adjusted returns. We find that when top-performing hedge funds are identified based on prior pooled alphas, the top decile portfolio of hedge funds earns a pooled alpha of 4.19% per year and an annualized Sharpe ratio of 1.05. By contrast, the top decile portfolios formed based on the Fung and Hsieh (2004) model, stepwise regression, and the style-adjusted method have pooled alphas of only 3.09%, 2.27%, and 2.50%, respectively, with Sharpe ratios of 0.82, 0.65, and 0.56. The performance differences are significant in both statistical and economic terms and also generally hold across various hedge fund investment styles.

Finally, we examine the value of pooled benchmarks in characterizing failure risk. We estimate a series of proportional hazards models to examine the failure rates for hedge funds classified ex ante as "poor" performers based on their ranking in the

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

lowest quartile of funds according to the (a) pooled alphas, (b) Fung and Hsieh (2004) model alphas, (c) stepwise regression model alphas, and (d) style-adjusted returns. We find that the marginal effect of being classified in the low pooled alpha group is a 66% increase in the baseline hazard. In comparison to the other three approaches, relying on pooled alphas also results in better identification of subsequent failures.

There have been a number of attempts to improve performance benchmarks for hedge funds following the important study by Fung and Hsieh (1997), which documents that hedge fund returns have low and sometimes negative correlations with traditional asset classes. Studies in this area have sought to expand the set of factors to be used in the benchmark model in order to capture the distinctive nature of hedge fund strategies. The factors are typically identified using statistical techniques like factor analysis (e.g., Fung and Hsieh 1997) or stepwise regression (e.g., Liang 1999, Fung and Hsieh 2000, Agarwal and Naik 2004, Bollen and Whaley 2009).[2] We extend this literature by allowing for the possibility that none of the benchmark models fully captures the investing style of a particular hedge fund. Accordingly, we consider a benchmark that combines or pools a diverse set of factors for hedge fund performance evaluation.

2. Methodology

The goal of investment performance evaluation is to assess the value added by an active fund manager relative to a passive benchmark. Following Sharpe's (1992) pioneering work, it is now standard practice to decompose the realized fund return into two components—a component that reflects the return that could be earned by suitable exposure to a set of passive benchmark (asset class) factors and a residual component that reflects managerial "skill." The challenge is to specify a benchmark or set of factors that will most closely mirror the investing "style" of the fund in terms of its exposure to systematic factors.[3]

[2] Recent studies have also attempted to account for potential changes in funds' factor exposures, which are identified using either an optimal changepoint detection framework (e.g., Bollen and Whaley 2009) or a structural break framework (e.g., Fung and Hsieh 2004). In related work, Patton and Ramadorai (2013) study the dynamics of hedge funds' factor exposures using high-frequency conditioning variables. Another strand of the literature on performance evaluation has sought to address the implications of the serial correlation in reported hedge fund returns for their performance statistics, based on an econometric model of smoothed returns (see, e.g., Getmansky et al. 2004).

[3] In this context Ferson (2010) formalizes the notion of an *appropriate benchmark* as one that has the same covariance with the relevant stochastic discount factor as the fund being evaluated. Ferson (2013) provides a detailed review of the various performance evaluation measures, including the conditional measures, employed in the literature.

In this section we describe the methodology for pooling candidate factor models, i.e., the strategy coherent factor subsets, with a view to constructing a pooled benchmark for assessing the performance of individual hedge funds. The factor subsets may be viewed as attribution models that help explain the individual hedge fund returns. The weight assigned to a given set of factors in the pooled benchmark reflects both the ability of a particular factor subset in explaining fund returns and its diversification benefits when combined with the other candidate factor models. Intuitively, the pooled benchmark is designed to best capture the investing style, and hence the realized returns, of the hedge fund strategy.

2.1. Model Implied Conditional Return Distributions

Consider a candidate factor model that describes the returns to the trading strategy of a particular hedge fund. Following standard practice, the hedge fund strategy is assumed to be characterized by the fund's exposure to the model's style factors. More formally, assume that the conditional expectation function for the fund excess return is linear in the factors and the disturbance vector. Then the fund's excess return at time t is

$$r_t = \alpha + \beta' f_t + \epsilon_t, \quad \epsilon_t \sim \text{i.i.d. } N(0, \sigma^2), \tag{1}$$

where r_t is the excess return at time t, f_t is a $k \times 1$ vector of k factors at time t, and ϵ_t is the disturbance at time t. Note that in the case of hedge funds the factors are often designed to mimic the nonlinear payoffs of dynamic hedge fund strategies. The parameter vector β has dimension $k \times 1$. For the purpose of estimation, it is convenient to specify the model in matrix notation. For a sample of length τ, the model is

$$R_{t-\tau,\,t-1} = F_{t-\tau,\,t-1} B + U_{t-\tau,\,t-1}, \tag{2}$$

where the subscript $t-\tau, t-1$ denotes a sample that extends from month $t-\tau$ to month $t-1$. In Equation (2) the ith element of the $\tau \times 1$ vector $R_{t-\tau,\,t-1}$ is $r_{t-\tau-1+i}$, the ith row of the $\tau \times (k+1)$ matrix $F_{t-\tau,\,t-1}$ is $(1, f'_{t-\tau-1+i})$, the ith element of the $\tau \times 1$ vector $U_{t-\tau,\,t-1}$ is $\epsilon'_{t-\tau-1+i}$, and $B = [\alpha, \beta']'$ is a $(k+1) \times 1$ vector of regression parameters. It is assumed that $\text{rank}(F_{t-\tau,\,t-1}) = k+1$. By the earlier assumption on ϵ_t, the elements of $U_{t-\tau,\,t-1}$ are i.i.d. $N(0, \sigma^2)$.

The focus in this research is on evaluating and pooling factor models based on their ability to explain hedge fund returns. The evaluation of the factor models is based on their respective conditional distributions. This study implements a Bayesian approach to obtain the conditional density for hedge fund returns.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

For simplicity, assume that the investor has the standard uninformative prior with respect to the fund's factor exposures, B:

$$p(B, \sigma^2 \mid R_{t-\tau,\,t-1}, F_{t-\tau,\,t-1}) \propto \frac{1}{\sigma^2}. \tag{3}$$

The posterior distribution of the parameters of interest is then given by

$$\sigma^2 \mid R_{t-\tau,\,t-1}, F_{t-\tau,\,t-1} \sim IG(\tau-(k+1), S), \tag{4}$$

$$B \mid R_{t-\tau,\,t-1}, F_{t-\tau,\,t-1}, \sigma^2 \sim N(\hat{B}, \sigma^2(F'_{t-\tau,\,t-1}F_{t-\tau,\,t-1})^{-1}), \tag{5}$$

where $S = (R_{t-\tau,\,t-1} - F_{t-\tau,\,t-1}\hat{B})'(R_{t-\tau,\,t-1} - F_{t-\tau,\,t-1}\hat{B})$, and $\hat{B} = [\hat{\alpha}, \hat{\beta}']'$ is the matrix of maximum likelihood (ML) parameter estimates.

Note that the marginal posterior probability density function (pdf) for σ^2 is in the form of the inverse gamma (IG) distribution, and the conditional posterior pdf for B is multivariate normal. Following Zellner (1971), the expression for the one-step-ahead conditional density for r_t has the form of the univariate student t distribution:

$$p(r_t \mid R_{t-\tau,\,t-1}, F_{t-\tau,\,t-1}, f_t) = \frac{\Gamma[(\nu+1)/2]}{\Gamma(1/2)\Gamma(\nu/2)}\left(\frac{h}{\nu}\right)^{1/2} \cdot \left[1+\frac{h}{\nu}(r_t-\hat{\alpha}-\hat{\beta}'f_t)^2\right]^{-(\nu+1)/2}, \tag{6}$$

where Γ denotes the gamma function, f_t is the vector of one-step-ahead factor realizations, $h = g\nu/(r_t - \hat{\alpha}-\hat{\beta}'f_t)^2$, $g = 1-f_t'(\bar{F}'\bar{F}+f_t'f_t)^{-1}f_t$, $\bar{F} = \{f_{t-\tau}, \ldots, f_{t-1}\}$, and the degrees of freedom $\nu = \tau-(k+1)$.

To determine the appropriate weights of the candidate models in the pool, models are evaluated based on their respective out-of-sample performance. For the purpose of model evaluation, $\hat{\alpha}$, which has the connotation of manager skill, is set equal to zero in the above conditional density. We note that this convention is suitable in the present context since we are interested in constructing a pooled passive benchmark that best describes the returns to the (unobserved) investment strategy pursued by a hedge fund. The passive benchmark, by definition, is expected to have zero alpha.

2.2. Optimal Pooled Benchmark

We now describe our approach to combining performance attribution models to generate fund-specific optimal pooled benchmarks. We specifically adopt the model pooling framework in Geweke and Amisano (2011) and O'Doherty et al. (2012).[4] The approach relies on the log scoring rule to evaluate linear combinations of probability distributions implied by individual factor models. An optimal pool of models is one in which the individual model weights are chosen to maximize the log score function. There are a number of reasons that justify the use of the log score criterion to assess model performance in the present context. The log score is a well-known information-theoretic measure of the conditional prediction performance track record of a model.[5] In fact, as noted by Gelman et al. (2014), the log score is perhaps the most commonly used scoring rule for the purpose of model evaluation and selection.

Intuitively, the log score for a particular model reflects the degree of correspondence, or lack of divergence, between the conditional return distribution implied by the model and the realized return distribution. Thus, a model that attaches a high probability to the return outcome that is subsequently realized would receive a relatively high score. Interestingly, the use of the log score rule typically results in several of the models in the pool under consideration receiving positive weights, a desirable feature in contexts where the model space is incomplete in the sense that it does not include the "correct" model.

Calculation of the log score relies on the use of the model-implied conditional densities and the realizations of the time series r_t and f_t, where the latter are denoted by r_t^o and f_t^o (o for observed). For a sample consisting of r_t^o, f_t^o, $R^o_{t-\tau,\,t-1}$, and $F^o_{t-\tau,\,t-1}$, the log score function of a single model A is

$$LS(r_t^o, f_t^o, R^o_{t-\tau,\,t-1}, F^o_{t-\tau,\,t-1}, A) = \sum_{t=\tau+1}^{T} \log[p(r_t^o \mid R^o_{t-\tau,\,t-1}, F^o_{t-\tau,\,t-1}, f_t^o, A)]. \tag{7}$$

As mentioned above, and as is evident from Equation (7), the log score function is intuitively appealing as it gives a high score to the model that ex ante

[4] Our approach is related to the forecast combination methodology pioneered by Bates and Granger (1969). Recent surveys of this literature include Clemen (1989), Diebold and Lopez (1996), Newbold and Harvey (2002), and Timmermann (2006). Applications of the methodology include the use of combination forecasts for predicting macroeconomic variables (Stock and Watson 2003, Stock and Watson 2004, Guidolin and Timmermann 2009) and the equity premium (Rapach et al. 2010). In contrast to the earlier work, which focuses on point forecasts, our model pooling framework emphasizes combining conditional densities.

[5] For example, Klein and Brown (1984) propose a model selection criterion related to the log score. More generally, scoring rules are a useful tool in evaluating probability forecasts. The log score rule was introduced by Good (1952), and the connection between the log score rule and the Kullback–Leibler divergence is studied in Hall and Mitchell (2007). For a more extensive discussion of the link between scoring rules and expected utility maximization based decision rules, see Jose et al. (2008).

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

assigns a high conditional probability to the value of r_t that materializes, that is, r_t^o. As such, the log score function reflects the out-of-sample prediction performance of a given factor model conditional on the factor realizations. This is a particularly desirable feature of the methodology in the present context as the focus on out-of-sample model performance helps guard against overfitting.

This paper considers n alternative performance attribution models $A_1, \ldots, A_n$ for r_t conditional on f_t. The log scoring rule is used to evaluate linear combinations of the densities of the form

$$\sum_{i=1}^{n} w_i p(r_t^o \mid R^o_{t-\tau,\,t-1}, F^o_{t-\tau,\,t-1}, f_t^o, A_i); \qquad \sum_{i=1}^{n} w_i = 1; \quad w_i \geq 0 \quad (i = 1, \ldots, n). \tag{8}$$

The restrictions on the weights w_i are necessary and sufficient to assure that the linear combination of density functions is a density function for all values of the weights and all arguments of any density function. Note that in our context the conditional density produced by each model as shown in Equation (6) is heavy tailed. The combination of these densities is also heavy tailed, a desirable feature when modeling hedge fund returns. Also note that the model implied predictive densities for the one-step-ahead return, r_t, are conditional on the vector of the contemporaneous factor realizations, f_t. This is appropriate in a performance evaluation context in which the focus is on isolating the component of the realized fund return that is explained by exposure to factors specified by a particular model.

In the full sample design case, the log score function for a pool with n models is

$$\tilde{f}_T(\mathbf{w}) = \sum_{t=\tau+1}^{T} \log\left[\sum_{i=1}^{n} w_i p(r_t^o \mid R^o_{t-\tau,\,t-1}, F^o_{t-\tau,\,t-1}, f_t^o, A_i)\right], \tag{9}$$

where $\mathbf{w} = (w_1, \ldots, w_n)'$, $w_i \geq 0$ for $i = 1, \ldots, n$ and $\sum_{i=1}^{n} w_i = 1$. An optimal pool is one where the weights are chosen to maximize $\tilde{f}_T(\mathbf{w})$ subject to the restrictions noted above. Accordingly, the optimal pool corresponds to $\mathbf{w}_T^* = \arg\max_{\mathbf{w}} \tilde{f}_T(\mathbf{w})$.

The optimal weights for the individual benchmark models derived above are used to construct a weighted benchmark. Accordingly, the pooled benchmark return for period t is given by $\sum_{i=1}^{n} w_i^* \hat{\beta}'_{i,t} f_{i,t}$, where w_i^* denotes the optimal weight assigned to factor model i in the pool. Hedge fund (pooled) alphas are then calculated with respect to the pooled benchmark constructed in this fashion.

2.3. Model Weights and Empirical Design Issues

The conditional densities for the asset pricing models use a three-year rolling window of monthly time-series data, that is, $\tau = 36$ to calculate $\hat{\beta}$ and S. Accordingly, for a hedge fund with return data available from time $t = 1$ to $t = T$, the first density is calculated for date $t = 37$. Note, therefore, that throughout the paper we are focusing on conditional versions of the performance attribution models that allow for time variation in the estimated factor loadings to account for changing fund investment styles. In this sense the pooled alphas reported in the paper should be interpreted as conditional performance measures.

In the empirical applications that follow, the model weights for a given fund are obtained for two different designs: the full sample and rolling two-year windows. In the case of the full sample design, the log score function value is calculated for date T using monthly conditional densities from date $t = 37$ through $t = T$, where T is the final month of the time series. For the two-year rolling window design, the first log score function value is calculated using conditional densities from date $t = 37$ to date $t = 60$. The log score values for subsequent months are calculated using only the previous two years of monthly conditional densities. Modification to Equation (9) to accommodate rolling window estimation is straightforward.

3. Data

Section 3.1 describes the sample of hedge funds used in the empirical analysis. Section 3.2 outlines the set of risk factors and performance attribution models used in the paper.

3.1. Hedge Funds

We obtain data on individual hedge funds from the Lipper TASS database for the period from January 1994 to December 2011. For each sample fund, we collect the history of monthly returns, the reported trading strategy (PrimaryCategory), the currency of reported returns (CurrencyCode), and information about the fund's use of a high-water mark provision (HighWaterMark), personal investment by managers (PersonalCapital), leverage (Leveraged), and lockup provision (LockUpPeriod). For funds that are dropped from the database prior to December 2011, we also obtain the reported reason for the end to the fund's reporting history (DropReasonid). We convert all returns to excess returns by subtracting the monthly risk-free rate.[6]

[6] Data on the risk-free rate are from Kenneth French's website (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/, accessed July 6, 2012). We thank Kenneth French for making these data available.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 1 **Summary Statistics for Monthly Excess Returns of Lipper TASS Funds**

	Number of funds			Distributional statistics			
Category	Live	Defunct	Total	Mean	Std. dev.	Skewness	Kurtosis
Panel A: Full sample							
All Funds	1,405	1,628	3,033	0.41	3.79	−0.51	5.36
Panel B: By fund category							
Directional funds							
Dedicated short bias	6	16	22	−0.12	6.51	0.19	3.71
Emerging markets	119	90	209	0.70	5.93	−0.60	5.95
Global macro	41	59	100	0.52	4.35	0.26	2.89
Managed futures	130	114	244	0.51	5.89	0.24	2.58
Nondirectional funds							
Convertible arbitrage	26	57	83	0.34	2.81	−1.29	12.02
Equity market neutral	33	66	99	0.35	3.04	−0.42	8.21
Fixed income arbitrage	24	58	82	0.32	2.76	−1.78	17.05
Semidirectional funds							
Event driven	105	149	254	0.42	2.69	−0.73	6.01
Long/short equity hedge	362	451	813	0.59	4.78	−0.04	3.60
Multistrategy	75	75	150	0.36	3.20	−0.73	7.39
Fund of funds							
Fund of funds	484	493	977	0.21	2.49	−0.92	5.38

Notes. Panel A reports summary statistics for all sample hedge funds, and panel B reports statistics for funds by category. The summary statistics are the average monthly excess return, standard deviation, skewness, and excess kurtosis. Within a given category, the figures presented are the equally weighted averages of the statistics across sample funds. The sample includes funds in the Lipper TASS database with at least 60 months of consecutive return data, monthly net-of-fee reported returns, and a currency code of "USD" (U.S. dollar). The sample period is from January 1994 to December 2011.

In constructing the sample, we require funds to have dollar returns (i.e., CurrencyCode of "USD") and exclude funds that only provide gross-of-fee returns or only report returns on a quarterly basis. We impose a filter to mitigate the impact of backfill bias on our results by eliminating the first 24 months of returns for each fund.[7] We also require funds to have one of the following 11 reported trading strategies: convertible arbitrage, dedicated short bias, emerging markets, equity market neutral, event driven, fixed income arbitrage, fund of funds, global macro, long/short equity hedge, managed futures, and multistrategy. For many of our empirical tests, we further group sample funds into one of four style categories following the classification scheme in Bali et al. (2012): directional funds (i.e., dedicated short bias, emerging markets, global macro, and managed futures), nondirectional funds (convertible arbitrage, equity market neutral, and fixed income arbitrage), semidirectional funds (event driven, long/short equity hedge, and multistrategy), and fund of funds.

As of December 2011, there are 8,112 hedge funds that satisfy the data requirements detailed above. Of these funds, 2,746 are live and 5,366 are defunct. Finally, the log score algorithm described in §2 requires a sufficiently long return series to compute predictive densities and model weights for a given fund. As such, most of our empirical tests also require a fund to have at least 60 months of consecutive returns to be included in the analysis. This final sample contains 3,033 funds, of which 1,405 are live and 1,628 are defunct.

Panel A of Table 1 reports summary statistics for sample funds. The table lists the number of funds and the cross-sectional average of each fund's average monthly excess return, standard deviation, skewness, and excess kurtosis. The average fund has a mean excess return of 0.41% per month with a standard deviation of 3.79%. Sample funds also tend to have negatively skewed returns and exhibit thick tails in their return distributions. Panel B presents summary statistics for each of the individual hedge fund trading strategies. The three most common categories are fund of funds (977 funds), long/short equity hedge (813), and event driven (254). Fund performance differs widely across categories. For example, the average dedicated short bias fund has a mean excess return of −0.12% per month, whereas emerging markets funds have an average excess return of 0.70%. All fund categories show evidence of excess kurtosis, but these figures are most pronounced for the

[7] Backfill bias arises because funds typically report to the Lipper TASS database only after they have been in operation for some period of time. When a fund enters the database, the manager is also able to report a past history of returns. This scheme leads to a bias in reported hedge fund performance because funds are less likely to backfill following periods of poor returns. We also control for biases that would arise from hedge funds being late in reporting to the Lipper TASS database. We use data only through December 2011 from a version of the database obtained in July 2012.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 2 Summary Statistics for Factors

Factor	Description	Mean	Std. dev.	Skewness	Kurtosis
	Panel A: Domestic equity factors				
MKTXS	Excess return of the CRSP value-weighted index	0.48	4.71	−0.72	1.20
SMB	Fama–French size factor	0.21	3.60	0.85	7.93
HML	Fama–French value factor	0.23	3.42	0.03	2.78
UMD	Momentum factor	0.49	5.51	−1.59	9.93
LIQ	Pastor–Stambaugh liquidity factor	0.76	4.05	0.57	2.88
	Panel B: Fixed income and commodity factors				
D10YR	Barclays 7–10 Treasury Index minus risk-free rate	0.32	1.88	0.12	1.24
DSPRD	Barclays Baa Index minus Barclays 7–10 Treasury Index	0.00	1.76	−0.39	8.03
GSCI	S&P GSCI commodity index	0.34	6.48	−0.38	1.40
	Panel C: Global factors				
EXUSA	MSCI world excluding United States	0.06	4.90	−0.62	1.49
EM	MSCI emerging markets	0.24	7.09	−0.71	1.78
DOL	Trade weighted U.S. dollar index	−0.35	1.69	0.12	0.97
	Panel D: Trend following factors				
PTFSBD	Primitive trend follower strategy bond	−1.24	15.26	1.35	2.50
PTFSFX	Primitive trend follower strategy currency	−0.18	19.58	1.36	2.65
PTFSCOM	Primitive trend follower strategy commodity	−0.38	13.65	1.18	2.37
PTFSIR	Primitive trend follower strategy interest rate	1.35	27.09	4.30	26.73
PTFSSTK	Primitive trend follower strategy stock	−4.98	13.70	1.41	3.81
	Panel E: Index option factors				
ATMC	S&P 500 at-the-money call	−3.10	77.21	0.58	−0.65
OTMC	S&P 500 out-of-the-money call	−4.76	82.63	0.73	−0.45
ATMP	S&P 500 at-the-money put	−18.39	90.14	1.75	3.07
OTMP	S&P 500 out-of-the-money put	−20.84	92.62	1.90	3.72

Notes. The table lists the performance attribution models included in this paper and reports the following summary statistics for the associated factor returns: average monthly return, standard deviation, skewness, and excess kurtosis. The performance attribution models considered are the domestic equity five-factor model (panel A), the fixed income and commodity three-factor model (panel B), the global three-factor model (panel C), the trend following five-factor model (panel D), and the index option four-factor model (panel E). The factors are described in the text. The sample period for all factor returns is from January 1994 to December 2011.

nondirectional funds. These results are likely related to differences in asset liquidity and leverage across hedge fund styles. In this spirit, Ang et al. (2011) show that convertible arbitrage and fixed income arbitrage funds exhibit relatively high levels of gross leverage, which may amplify profits and losses for many of these funds.

3.2. Factors and Models

The optimal pooled benchmarks considered in this paper are formed by combining five performance attribution models, each of which can be viewed as a strategy-coherent subset of style-based factors. These models and their associated factors are listed in Table 2. The first model incorporates five domestic equity factors that are widely used in the literature on modeling the performance of managed portfolios. This group includes the excess return on the CRSP value-weighted stock market index (*MKTXS*), the size (*SMB*) and value (*HML*) factors from the Fama and French (1993) three-factor model, the momentum factor (*UMD*) designed to capture the difference in performance between stocks that have performed well and poorly over the prior year, and the traded liquidity factor (*LIQ*) from Pastor and Stambaugh (2003).[8]

The second model combines two bond market factors with a commodity markets index. The return on the Barclays Capital 7–10 year Treasury Index in excess of the risk-free rate (*D10YR*) and the return spread of the Barclays Corporate Bond Baa Index minus the 7–10 year Treasury Index (*DSPRD*) are used to measure fund exposure to the domestic fixed income market. We add the excess return for the S&P GSCI (*GSCI*) to benchmark for hedge fund investments in commodity markets.

The third model included in the pool is a three-factor global model intended to proxy for potential hedge fund exposure to foreign equity and foreign exchange markets. The international equity factors are the return for the Morgan Stanley Capital International World Index ex United States (*EXUSA*) and the return for the Morgan Stanley Capital International

[8] The market, size, value, and momentum factors are from Kenneth French's website. The liquidity factor is available through Wharton Research Data Services.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Emerging Markets Index (*EM*).[9] The returns for both of these series are converted to excess returns by subtracting the risk-free rate. We also use the return on the trade weighted U.S. dollar index in excess of the risk-free rate (*DOL*) to account for fund exposure to foreign exchange markets.[10]

The fourth and fifth models consist of factors explicitly designed to model hedge fund returns. The Fung and Hsieh (2001) five-factor model includes the Primitive Trend Following Strategies for bonds (*PTFSBD*), exchange rates (*PTFSFX*), commodities (*PTFSCOM*), interest rates (*PTFSIR*), and stocks (*PTFSSTK*). These factors are excess returns on portfolios of look-back straddle options on the related asset styles.[11] Finally, the index option model further accounts for fund exposure to nonlinear factors by including the Agarwal and Naik (2004) portfolios of at-the-money calls (*ATMC*), out-of-the-money calls (*OTMC*), at-the-money puts (*ATMP*), and out-of-the-money puts (*OTMP*) on the S&P 500 Index.[12]

Table 2 presents summary statistics for the individual factors. There is substantial variation in mean excess returns and standard deviations. The five trend following factors and the four portfolios of call and put options have the highest risk measures with standard deviations ranging from 13.65% to 92.62% per month. Of these nine nonlinear factors, eight also have negative average returns. The remaining nonlinear factor, *PTFSIR*, exhibits high levels of skewness and kurtosis over the sample period, driven largely by extreme returns in August 2007 (quant crisis) and from September to October 2008 (peak of the recent global financial crisis).

The objective of the paper is to demonstrate the benefits of incorporating information from several competing models in characterizing hedge fund performance. Our choice of defining the constituent models as five strategy-coherent subsets of factors also yields a convenient interpretation of our modeling procedure as one with two steps. In the first step, we measure the relation between a given fund's returns series and each of the five strategy-coherent models. In the second step, we assign weights to each strategy (i.e., model) via maximization of the log score objective function in Equation (9).

4. Results

Section 4.1 illustrates the statistical properties of pooled benchmarks. Section 4.2 compares the performance of these benchmark models to the Fung and Hsieh (2004) seven-factor model and stepwise regression approaches in replicating individual hedge fund returns. Section 4.3 discusses survivorship issues with the sample and considers the ability of our pooling algorithm to characterize portfolios based on fund age. Sections 4.4 and 4.5 explore the economic benefits to model combination in forecasting hedge fund returns and failures, respectively.

4.1. Statistical Properties of Optimal Pooled Benchmarks

As a starting point, we consider the optimal five-model pool for each sample hedge fund using the entire history of returns. We estimate the model weights for a given fund following Equation (9) and require a minimum of two years of conditional densities to be included in the sample. Table 3 reports summary statistics for the pools. Panel A presents the average weight assigned to each of the five models across all hedge funds and across each fund category. The largest average weight across all hedge funds is assigned to the three-factor global model (31%), suggesting a substantial exposure to foreign equity and foreign exchange markets within our sample of funds. The domestic equity model and the fixed income and commodity model are also assigned relatively large average weights at 25% each. Note, however, that all five of the models are allocated a meaningful weight. The smallest average weight is assigned to the index option model at 9%.

The results for the individual fund categories listed in panel A also suggest that the log score algorithm does a satisfactory job of assigning high weights to models that we would expect ex ante to be important within each group. For example, the average weight assigned to the three-factor global model is relatively high among funds in the emerging markets (61%) and global macro groups (31%). The managed futures funds give a high average weight (32%) to the set of Fung and Hsieh (2001) trend following factors that were proposed specifically to capture the strategies of commodity trading advisors (CTAs). Similarly, funds reporting to follow the dedicated short bias and long/short equity strategies have the largest average exposure (63% and 44%, respectively) to the domestic equity model. Thus, the log score maximization approach appears to yield an economically meaningful characterization of hedge fund performance.

[9] The data for these factors are available at http://www.msci.com (accessed June 23, 2013).

[10] The U.S. dollar index series is from the Federal Reserve Bank of St. Louis website (http://research.stlouisfed.org/fred2/, accessed December 14, 2012).

[11] The trend following factors are from David Hsieh's website (http://faculty.fuqua.duke.edu/~dah7/HFData.htm, accessed December 9, 2012). We thank David Hsieh for making these data available.

[12] Vikas Agarwal generously provided these data for the period from January 1994 through August 2009. We construct factor returns for the remainder of the sample period following the method described in Agarwal and Naik (2004) using options data provided by the Options Price Reporting Authority (OPRA).

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 3 Summary Statistics for Optimal Five-Model Pools

	Model					Number of models	
Category	Domestic equity	Fixed income and commodity	Global	Trend following	Index option	Mean	Median
			Panel A: Average model weights and number of models				
All funds	0.25	0.25	0.31	0.11	0.09	2.87	3.00
Directional funds							
Dedicated short bias	0.63	0.03	0.15	0.07	0.12	2.36	2.00
Emerging markets	0.10	0.19	0.61	0.05	0.06	2.47	2.00
Global macro	0.14	0.27	0.31	0.18	0.10	2.86	3.00
Managed futures	0.08	0.31	0.19	0.32	0.09	2.90	3.00
Nondirectional funds							
Convertible arbitrage	0.18	0.51	0.13	0.07	0.11	3.06	3.00
Equity market neutral	0.30	0.27	0.23	0.10	0.10	2.92	3.00
Fixed income arbitrage	0.19	0.35	0.17	0.20	0.10	3.17	3.00
Semidirectional funds							
Event driven	0.27	0.30	0.19	0.10	0.14	3.19	3.00
Long/short equity hedge	0.44	0.14	0.28	0.05	0.09	2.72	3.00
Multistrategy	0.16	0.33	0.25	0.13	0.13	3.03	3.00
Fund of funds							
Fund of funds	0.18	0.27	0.38	0.10	0.06	2.92	3.00
			Panel B: Fraction of funds with positive model weight				
All funds	0.61	0.66	0.74	0.45	0.41		
Directional funds							
Dedicated short bias	0.95	0.23	0.50	0.23	0.45		
Emerging markets	0.42	0.57	0.87	0.31	0.31		
Global macro	0.43	0.66	0.76	0.57	0.44		
Managed futures	0.35	0.76	0.64	0.77	0.38		
Nondirectional funds							
Convertible arbitrage	0.67	0.92	0.49	0.46	0.52		
Equity market neutral	0.73	0.67	0.67	0.45	0.40		
Fixed income arbitrage	0.66	0.80	0.57	0.65	0.49		
Semidirectional funds							
Event driven	0.72	0.79	0.66	0.49	0.52		
Long/short equity hedge	0.78	0.48	0.72	0.29	0.45		
Multistrategy	0.52	0.77	0.61	0.56	0.57		
Fund of funds							
Fund of funds	0.54	0.73	0.83	0.48	0.34		

Notes. The table reports summary statistics for optimal five-model pools. The performance attribution models considered are the domestic equity five-factor model, the fixed income and commodity three-factor model, the global three-factor model, the trend following five-factor model, and the index option four-factor model. For a given fund, the log score and optimal weights are computed from conditional densities based on the maximization of the log score objective function using the entire sample of fund returns. Panel A reports average model weights and the mean and median number of models assigned a positive weight across all sample funds and by fund category. Panel B reports the proportion of funds with positive weight on each model.

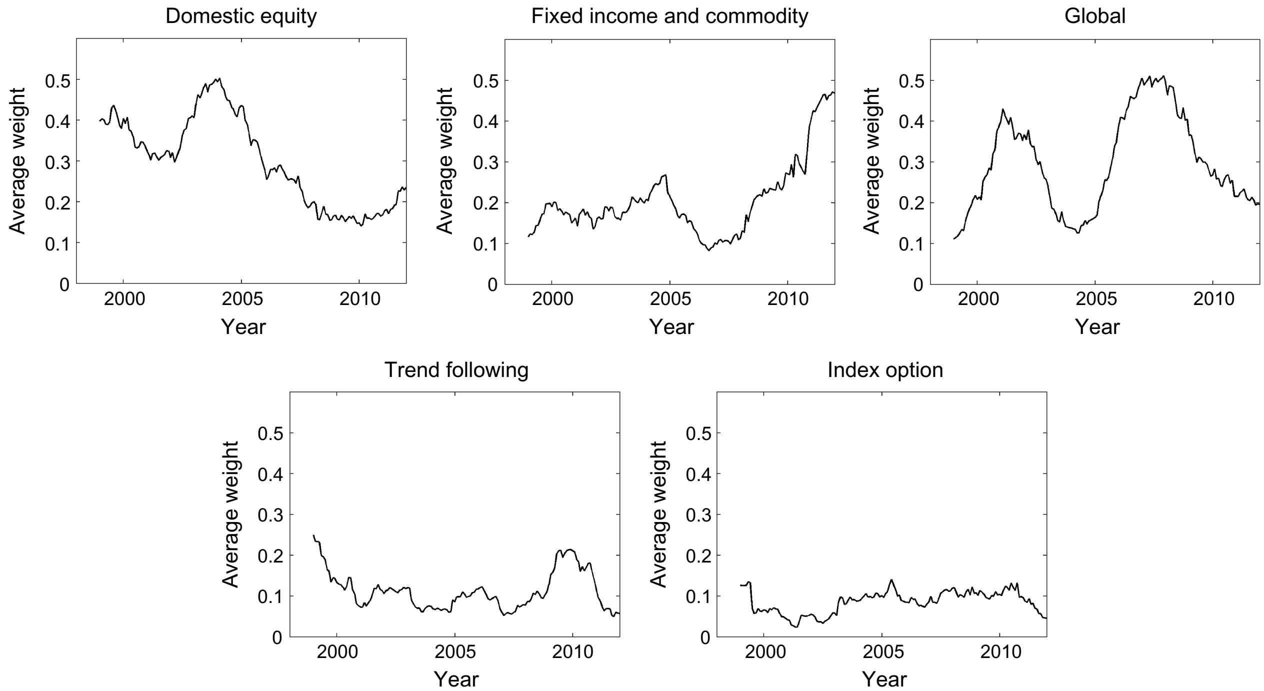
Panel A of Table 3 also reports the mean and median number of models included across all sample funds and for each of the fund categories. For the full sample of 3,033 funds, the average number of models selected is 2.87 and the median is three. These results are similar across the individual classifications. The mean number of models ranges from 2.36 for dedicated short bias funds to 3.19 for event driven funds. In unreported results, we find that the log score algorithm identifies a single model as sufficient in only 7.3% of sample funds. A large percentage of funds (64.8%) assign positive weight to at least three models.

Panel B of Table 3 shows the fraction of funds that assign positive weight to each model. The global model is included in 74% of the sample pools. The next most popular model is the fixed income and commodity model (66%), followed by the domestic equity model (61%). The least popular model is the index option model, which is included only 41% of the time. The most interesting feature of these results is that none of the models is consistently included in the pools across all sample funds. For example, although the global three-factor model is the most commonly included model, it is allocated a weight of zero in over one quarter of the optimal pools. For

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Figure 1 **Evolution of Average Model Weights with Rolling Windows in the Five-Model Pools**



Notes. The figure shows the evolution of average model weights across all sample funds for the optimal five-model pools. The weights for a given fund are updated each month using the prior two years of conditional densities. The performance attribution models considered in the optimal pools are the domestic equity five-factor model, the fixed income and commodity three-factor model, the global three-factor model, the trend following five-factor model, and the index option four-factor model.

these funds, the global model exhibits zero incremental explanatory power for fund returns in the presence of the remaining four performance attribution models. Perhaps more surprisingly, this result also extends to each of the fund categories. The trend following model, for example, which was originally designed to model the performance of managed futures funds, is completely excluded from 23% of the pools within this category. Thus, it is easy to see the potential pitfalls of relying on any individual model in characterizing the cross section of hedge fund risk exposures and quantifying managerial performance.

The figures reported in Table 3 are based on a constant set of model weights for each fund. We would also like to document the extent to which the relative weight allocated to each of the individual models changes over the sample period. We proceed by estimating model weights for each sample fund on a rolling basis. For each fund month we maximize the log score objective function in Equation (9) using the prior two years of predictive densities. Figure 1 reports the time series of average model weights across sample funds. The plots reveal moderate instability in the weights assigned to each of the models over time. Notably, the domestic equity model achieves an average weight of 50.0% in 2003, which generally declines over the remainder of the sample period. In contrast, the three-factor global model initially receives a modest weight of 11.1% in 1999 but achieves substantially more weight in recent years, indicating greater exposure to international equity and foreign exchange markets for sample funds. The fixed income and commodity and trend following models also see a spike in relative importance during the recent financial crisis. We note, however, that each of the five models in the pools is assigned positive and economically meaningful average weight throughout the sample period.

4.2. Hedge Fund Tracking

Having established the properties of the optimal five-model pools, we next turn to a statistical comparison of model pooling to other performance attribution approaches applied in the hedge fund literature. In particular we compare the tracking performance of the optimal pool for a given fund to that of either the Fung and Hsieh (2004) seven-factor model or a stepwise regression model. The seven-factor model incorporates a broad subset of the factors listed in Table 2: *MKTXS*, *SMB*, three of the trend following factors (*PTFSBD*, *PTFSFX*, and *PTFSCOM*), and the two bond market factors (*D10YR* and *DSPRD*). As mentioned previously, stepwise regression methods and other variable selection techniques are also popular alternatives to address model uncertainty by

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

customizing benchmarks to funds with potentially unique factor exposures and investment styles.

We focus on whether the optimal five-model pool for each fund produces significantly more accurate forecasts than either the Fung and Hsieh (2004) model or a stepwise regression model conditional on the realization of the factor returns. This is equivalent to comparing the accuracy of the competing density forecasts or the ability of each model to track the time-series variation in a given fund's returns. We specifically follow the formal statistical test outlined by Giacomini and White (2006) and Amisano and Giacomini (2007).

Amisano and Giacomini (2007) introduce a likelihood ratio test of the null hypothesis that two competing density forecasts, $\hat{f}$ and $\hat{g}$, are equally accurate. In the present context, we have candidate conditional densities for each sample fund implied by either the optimal five-model pool or an individual model. These densities are formed at date $t-1$ for the one-period-ahead fund excess returns, r_t. For a given hedge fund, the test is based on the difference in average log scores for the conditional densities for the fund's excess returns computed over the out-of-sample period:

$$\overline{WLR} = \frac{1}{T-(\tau+d)} \sum_{t=\tau+d+1}^{T} [\log \hat{f}(r_t) - \log \hat{g}(r_t)], \quad (10)$$

where $T-(\tau+d)$ denotes the number of out-of-sample forecasts being evaluated, $\tau = 36$ is the length of the rolling window of data used to compute factor loadings and construct the conditional densities, and $d = 24$ denotes the length of the rolling window of conditional densities used to compute model weights prior to each forecast month. The test is based on the statistic

$$t = \frac{\overline{WLR}}{\hat{\sigma}/\sqrt{T-(\tau+d)}}, \quad (11)$$

where $\hat{\sigma}^2$ is an estimator of the asymptotic variance, $\sigma^2 = Var(\sqrt{T-(\tau+d)}\ \overline{WLR})$. Under the null hypothesis of equal accuracy of the conditional densities, the above statistic is asymptotically distributed as a standard normal variate.

Before discussing the results, we first describe how we compute the conditional densities for each alternative. For the Fung and Hsieh (2004) model, the density for a given fund month is simply provided by Equation (6), where the factor loadings in each month are estimated using realized fund returns over the prior 36 months. For the optimal pool, the density is given by Equation (8), where the weights for a given fund month are estimated by maximizing the objective function in Equation (9) using conditional densities from the previous 24 months.[13] For the stepwise regression approach, we estimate a density for each fund in each month via a stepwise regression algorithm that uses the prior 36 months of returns data. For a given fund month, the algorithm chooses a subset of the 20 factors listed in Table 2 according to a stepwise regression process with an entry significance level of 15% and a retaining significance level of 50%.[14] Note that for a given fund, the stepwise model is updated each month so that the identity of the factors and the estimated loadings are allowed to change over time. This approach allows for a fair comparison to the pooled densities that are also based on factors estimated on a rolling 36-month basis. Once the identity of the factors is determined for a given fund month, the stepwise model density is also given by Equation (6).

Table 4 summarizes the formal statistical tests of the null, H_0: $\overline{WLR} = 0$, against the alternative, $\overline{WLR} > 0$, for cases in which the performance of the five-model pool is compared to that of either the Fung and Hsieh (2004) model or the stepwise regression approach. The table reports the fraction of sample funds for which $\overline{WLR}$ is greater than or equal to zero, i.e., the optimal pool outperforms the competing model, and the fraction of funds for which the difference is significantly greater than zero at the 10%, 5%, and 1% levels. These figures are provided for all funds and for the various fund classifications. We require a minimum out-of-sample period of 24 months for a given fund to be included in the results.

The results in Table 4 suggest that the optimal five-model pool performs at least as well as the Fung and Hsieh (2004) model (stepwise model) in 88% (96%) of sample funds. These differences in model performance are often statistically significant. The pool outperforms the seven-factor model at the 10% level in 52% of sample funds and at the 5% level in 40% of sample funds. For the stepwise model, the differences are statistically significant at the 10% (5%) level in 75% (61%) of funds. These general conclusions also hold for each of the four broadly defined fund categories.

Although the results in Table 4 provide strong statistical support for model pooling based on the log predictive score criterion, we would also like to characterize the extent to which model pooling results

[13] The pools summarized in Table 3 are based on a constant set of weights for each fund. In Table 4, we examine the performance of pools with the conditional densities on a given date based on weights estimated only from prior data and, therefore, avoid any concerns over look-ahead bias.

[14] We also require each fund to have at least one factor. Thus, if the stepwise algorithm suggests only a constant should be included in the regression model, we defer to the one-factor model with the highest R^2.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 4 Comparison of Log Scores

	Fung and Hsieh (2004) model				Stepwise regression model			
		Significance level				Significance level		
Category	$LS_{\text{Pool}} \geq LS_{FH7}$	10%	5%	1%	$LS_{\text{Pool}} \geq LS_{SW}$	10%	5%	1%
All funds	0.88	0.52	0.40	0.24	0.96	0.75	0.61	0.33
Directional funds	0.82	0.46	0.36	0.21	0.96	0.73	0.60	0.29
Nondirectional funds	0.81	0.30	0.19	0.08	0.98	0.77	0.66	0.44
Semidirectional funds	0.88	0.46	0.32	0.17	0.95	0.75	0.61	0.31
Fund of funds	0.93	0.67	0.57	0.38	0.97	0.76	0.62	0.34

Notes. The table compares log scores for optimal five-model pools, the Fung and Hsieh (2004) seven-factor model (*FH*7), and the stepwise regression model (*SW*). The performance attribution models considered in the model pools are the domestic equity five-factor model, the fixed income and commodity three-factor model, the global three-factor model, the trend following five-factor model, and the index option four-factor model. For a given fund, the optimal model weights are computed from conditional densities based on the maximization of the log score objective function using a two-year rolling window design. The conditional densities on a given date are therefore based on weights estimated from the prior two years of data. The table reports the proportion of funds for which the log score for the specified five-model pool (LS_{Pool}) is greater than the log score for either the Fung and Hsieh (2004) seven-factor model (LS_{FH7}) or the stepwise regression model (LS_{SW}) and the proportion for which the difference is significant at the 10%, 5%, and 1% levels. We require a minimum out-of-sample period of 24 months for a given fund to be included in the results.

in economically meaningful gains in tracking fund returns. We begin by comparing the ability of the individual factor models and the optimal pooled benchmark models to explain one-step-ahead returns for our sample of hedge funds. Following Fung and Hsieh (2001), we focus on the ability of each model to predict hedge fund returns conditional on estimated factor loadings and the factor realizations. Specifically, define the conditional forecast error for a given fund in month t relative to model i as $r_t - \hat{\beta}'_{i,t} f_{i,t}$, where $\hat{\beta}_{i,t}$ is a vector of factor loadings and $f_{i,t}$ is a vector of factor realizations during month t. For each sample hedge fund and each of the five constituent models, we compute the mean square forecast error across time. For a fund with T monthly return observations, the measure of the out-of-sample predictive performance for model i is thus

$$MSE_{\text{Model}} = \frac{1}{T-(\tau+d)} \sum_{t=\tau+d+1}^{T} (r_t - \hat{\beta}'_{i,t} f_{i,t})^2, \qquad (12)$$

where as before the factor loadings are estimated using a rolling 36-month window of fund returns prior to month t (i.e., $\tau = 36$).

It is straightforward to develop a corresponding performance measure for the five-model pool. We define the predicted excess return for a particular fund as the weighted average of the model excess return forecasts. The out-of-sample predictive ability of the pool is then given by

$$MSE_{\text{Pool}} = \frac{1}{T-(\tau+d)} \sum_{t=\tau+d+1}^{T} \Big(r_t - \sum_{i=1}^{n} w_{i,t} \hat{\beta}'_{i,t} f_{i,t}\Big)^2, \qquad (13)$$

where $w_{i,t}$ is the weight assigned to model i during month t. The model weights are allowed to vary over time, and the parameter d denotes the sample length of an initial set of predictive densities used to compute the first combination of model weights for each fund. In the results that follow, we use a rolling 24-month window of conditional densities (i.e., $d = 24$) to compute model weights prior to each forecast month. This empirical design leads to a genuine out-of-sample comparison, as the factor loadings for the individual models as well as the model weights are known prior to the excess return being forecasted.

For each fund and individual model we compute the difference in root mean square error to compare the ability of the individual model and the optimal pool to conditionally forecast returns. This measure is given by

$$\Delta RMSE = \sqrt{MSE_{\text{Model}}} - \sqrt{MSE_{\text{Pool}}}. \qquad (14)$$

This performance metric is straightforward to interpret as values of $\Delta RMSE$ greater than zero indicate that the optimal five-model pool for a given fund delivers superior conditional return forecasts over the out-of-sample period. Table 5 reports the cross-sectional average of the difference in root mean square error for all funds and for each fund classification. We require a minimum out-of-sample period length of 24 months for a fund to be included in the average.

Focusing first on the results in Table 5 for the individual constituent models, we see that the global and domestic equity models deliver the best performance relative to the optimal pool with average $\Delta RMSE$ values of 0.15% and 0.27% per month, respectively. The worst performing models are the index option and trend following models, with average performance scores of 0.77% and 0.79%, respectively. More importantly, the results indicate that, across all sample funds, the average $\Delta RMSE$ for the optimal pool relative to each of the five constituent models is positive.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 5 Comparison of Out-of-Sample Performance

Category	Constituent models: Domestic equity	Fixed income and commodity	Global	Trend following	Index option	Fung and Hsieh (2004) model	Stepwise model
All funds	0.27	0.45	0.15	0.79	0.77	0.41	0.44
Directional funds	0.59	0.55	0.07	0.95	1.14	0.57	0.60
Nondirectional funds	0.17	0.07	0.18	0.36	0.35	0.22	0.31
Semidirectional funds	0.18	0.59	0.22	1.05	0.84	0.41	0.46
Fund of funds	0.23	0.32	0.11	0.47	0.60	0.36	0.34

Notes. The table compares the out-of-sample replication performance of the optimal five-model pool to each of the five individual constituent models, the Fung and Hsieh (2004) seven-factor model, and the stepwise regression model. The table reports the average difference in root mean square error (in percentage per month) across sample funds for the optimal pool relative to each model. The difference in root mean square error is given by $\Delta RMSE = \sqrt{MSE_{\text{Model}}} - \sqrt{MSE_{\text{Pool}}}$, where MSE is the mean square prediction error. The prediction error for a given model and fund month is based on the fund's realized excess return during that month, the realization of the factor returns during that month, and the fund's risk exposures computed from the prior 36 months of fund returns. For the optimal pools, the prediction errors are also based on a rolling set of weights computed from the prior two years of conditional densities. The performance attribution models considered in the optimal pools are the domestic equity five-factor model, the fixed income and commodity three-factor model, the global three-factor model, the trend following five-factor model, and the index option four-factor model. Entries in the table that are greater than zero indicate that, on average, the optimal pool is outperforming the given individual model. The table shows averages across all funds and by fund category. We require a minimum out-of-sample period of 24 months for a given fund to be included in the averages.

Model pooling consistently leads to superior performance in tracking fund returns.

The model pooling approach also outperforms the Fung and Hsieh (2004) seven-factor model across all funds and within each of the four hedge fund classifications. The average $\Delta RMSE$ values range from 0.22% for nondirectional funds to 0.57% for the directional subcategory. For completeness, we consider the performance of the stepwise regression strategy described above in which the identity of the factors and the corresponding loadings are updated each month using the prior 36 months of data. The average $\Delta RMSE$ for the optimal pool relative to the stepwise regression method is 0.44% per month. Pooling also outperforms stepwise regression across each of the hedge fund subgroups.

4.3. Age Portfolios

One important limitation of the results to this point is that our sample is restricted to relatively established hedge funds. For example, the tests reported in Tables 4 and 5 require an initial 36 months of returns to estimate factor loadings for each constituent model, 24 months of data to estimate a fund's first set of model weights, and an additional 24 months to evaluate tracking performance. Given the short return histories and high attrition rates among hedge funds, it is important to address the statistical and economic benefits of model pooling in characterizing less established funds. We attempt to address this issue by adopting a portfolio approach. Specifically, we sort all sample funds into three age groups within each of the four broadly defined style categories.[15] "Young" hedge funds are those with less than two years of prior return data, "middle age" funds have between two and five years of prior return data, and "mature" funds have greater than five years of prior return data. The portfolios are rebalanced monthly, and funds are allowed to switch portfolios as they progress to the appropriate age.

Table 6 repeats much of the analysis in Tables 3–5 using our 12 age-sorted portfolios rather than individual hedge funds. Panel A reports optimal weights for each portfolio using the full sample of conditional densities. Within each style category, the estimated model weights are relatively similar across age portfolios. For example, the global three-factor model is allocated between 0.76% and 0.98% among the directional portfolios. The nondirectional portfolios assign substantial weight to the fixed income and commodity model, and each of the semidirectional and fund of funds portfolios gives considerable weights to the domestic equity and global models.

Panel B of Table 6 reports portfolio-by-portfolio results of likelihood ratio tests comparing the tracking ability of the optimal five-model pool to either the Fung and Hsieh (2004) model or stepwise model. In all cases the optimal pool exhibits superior tracking performance. As with the fund-level tests summarized in Table 4, these differences are often statistically significant. The optimal pool outperforms the seven-factor model at the 10% (5%) level in 11 (9) cases out of 12. Similarly, for the stepwise approach, the differences are significant at the 10% (5%) level for 10 (9) portfolios. Finally, panel C compares tracking performance based on differences in root mean square error. The $\Delta RMSE$ values for the pool in relation to the Fung and Hsieh (2004) model are greater than zero for all 12 portfolios. The pool outperforms

[15] In constructing the age portfolios, we follow the data cleaning procedure outlined in §3.1 with the exception of requiring funds to have 60 monthly returns.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 6 Age-Sorted Portfolios

	Panel A: Model weights					Panel B: Log score comparison		Panel C: $\Delta RMSE$ (%)	
Category	Domestic equity	Fixed income and commodity	Global	Trend following	Index option	Fung and Hsieh (2004) model	Stepwise model	Fung and Hsieh (2004) model	Stepwise model
Directional funds									
Young	0.00	0.02	0.98	0.00	0.00	0.000	0.001	0.63	0.27
Middle age	0.00	0.05	0.91	0.00	0.04	0.001	0.042	0.43	0.15
Mature	0.00	0.08	0.76	0.16	0.00	0.014	0.351	0.29	−0.07
Nondirectional funds									
Young	0.33	0.56	0.11	0.00	0.00	0.062	0.002	0.09	0.12
Middle age	0.10	0.69	0.22	0.00	0.00	0.065	0.005	0.13	0.17
Mature	0.27	0.50	0.14	0.00	0.09	0.335	0.045	0.08	0.07
Semidirectional funds									
Young	0.63	0.00	0.37	0.00	0.00	0.014	0.010	0.24	0.09
Middle age	0.69	0.00	0.31	0.00	0.00	0.003	0.132	0.31	0.09
Mature	0.80	0.00	0.20	0.00	0.00	0.023	0.061	0.22	0.09
Fund of funds									
Young	0.39	0.07	0.54	0.00	0.00	0.001	0.049	0.41	0.15
Middle age	0.29	0.09	0.61	0.00	0.00	0.000	0.006	0.39	0.22
Mature	0.34	0.13	0.53	0.00	0.00	0.002	0.028	0.30	0.11

Notes. The table repeats the analysis in Tables 3–5 using portfolios formed on hedge fund style and age rather than individual hedge funds. In a given style category, "young" funds have less than two years of prior return data, "middle age" funds have between two and five years of prior return data, and "mature" funds have greater than five years of prior return data. The portfolios are equally weighted and rebalanced monthly. Panel A reports the optimal model weights for each portfolio using the entire sample of portfolio returns. Panel B compares log scores for the optimal five-model pools, the Fung and Hsieh (2004) seven-factor model, and the stepwise regression model. For a given portfolio, the optimal model weights are computed from conditional densities based on the maximization of the log score objective function using a two-year rolling window design. The conditional densities on a given date are therefore based on weights estimated from the prior two years of data. The figures reported in panel B are p-values for the one-sided test that the log score for the specified five-model pool is greater than the log score for either the Fung and Hsieh (2004) seven-factor model or the stepwise regression model. Panel C compares the out-of-sample replication performance of the optimal five-model pool to the Fung and Hsieh (2004) seven-factor model and the stepwise regression model. The table reports the portfolio-level differences in root mean square error (in percentage per month) defined as $\Delta RMSE = \sqrt{MSE_{\text{Model}}} - \sqrt{MSE_{\text{Pool}}}$, where MSE is the mean square prediction error. The prediction error for a given model and portfolio month is based on the portfolio's realized excess return during that month, the realization of the factor returns during that month, and the portfolio's risk exposures computed from the prior 36 months of portfolio returns. For the optimal pools, the prediction errors are also based on a rolling set of weights computed from the prior two years of conditional densities.

the stepwise model in 11 out of 12 portfolios. Taken together, the results in Table 6 support the value of model pooling in characterizing the returns of hedge funds across age classifications.

4.4. Trading Strategies

Our results to this point show that model pooling is a valuable tool in capturing the risk exposures of individual funds and tracking one-step-ahead monthly returns conditional on past factor loadings and factor realizations. A potentially more relevant issue for investors is the extent to which pooling can be utilized to identify skilled fund managers. For example, Goetzmann et al. (2003) and Lan et al. (2013) show that the equilibrium required level of managerial alpha is large in order to justify the typical management and incentive fees imposed by hedge funds. Thus, investors' investment performance depends critically on their ability to assess managerial skill ex ante.

Prior research focusing on the issue of whether hedge fund performance persists has yielded mixed results.[16] In the context of hedge funds, however, there are theoretical reasons to believe that superior performance, if it exists, should persist through time. Glode and Green (2011), for example, propose a model in which the potential for information spillovers relating to profitable trading strategies leads to performance persistence. One potential concern with existing empirical studies on this issue is the reliance on individual models, likely specified with substantial error, to measure manager performance. We propose model pooling as an alternative to deliver a more refined measure of skill.

To investigate the value of model pooling in characterizing manager ability and predicting investment performance, we follow a portfolio approach. We contrast the performance of trading strategies based on

[16] For example, Agarwal and Naik (2000a, b), Baquero et al. (2005), Kosowski et al. (2007), and Jagannathan et al. (2010) find some evidence of performance persistence in their respective samples of hedge funds. Avramov et al. (2011) also report that fund alphas are predictable conditional on macroeconomic variables. In contrast, Brown et al. (1999), Capocci and Hubner (2004), and Fung et al. (2008) find little empirical support for persistence in hedge fund abnormal returns.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

prior alpha t-statistics from the optimal five-model pool, the Fung and Hsieh (2004) seven-factor model, the stepwise regression model, and a style-adjusted benchmark model, which simply compares the average return for a given fund to the equally weighted average return of all sample funds following the same trading strategy (e.g., convertible arbitrage, dedicated short bias, etc.).

The pooled alpha for a particular fund at the beginning of period $T+1$ is given by

$$\hat{\alpha} = \frac{1}{d} \sum_{t=T-d+1}^{T} \left(r_t - \sum_{i=1}^{n} w_{i,t} \hat{\beta}'_{i,t} f_{i,t} \right), \tag{15}$$

where the weights $w_{i,t}$ are computed from conditional densities spanning the period $t = T-d+1$ to $t = T$, the factor loadings $\hat{\beta}_{i,t}$ are estimated using a rolling 36-month window of fund returns prior to month t, and $d = 24$ is both the number conditional densities included in the estimation of the model weights and the number of monthly abnormal returns used to estimate alpha. In evaluating the ability of the optimal pools to predict hedge fund performance, we focus on portfolios sorted on the alpha t-statistics:

$$t(\hat{\alpha}) = \frac{\hat{\alpha}}{\hat{\sigma}_\alpha / \sqrt{d}}, \tag{16}$$

where $\hat{\sigma}^2_\alpha$ is an estimator of the asymptotic variance, $\sigma^2_\alpha = Var(\sqrt{d}\ \hat{\alpha})$. As Fama and French (2010) point out, t-statistics likely provide a more reliable indicator of managerial skill by accounting for differences in precision across alpha estimates. The alpha t-statistics for the Fung and Hsieh (2004) and stepwise trading strategies are computed analogously to Equations (15) and (16) with a weight of one allocated to the relevant model. The alpha t-statistics for the style-adjusted approach are calculated from the time series of fund returns in excess of the relevant style benchmark.

In January of every year, we first sort funds into decile portfolios within each hedge fund investment category (i.e., directional, nondirectional, semidirectional, and fund of funds) based on the t-statistics associated with fund alphas computed with respect to one of the four benchmark models. We then consider the performance of portfolios that invest in either all top decile funds or all bottom decile funds according to a given indicator. Our choice to first sort funds within categories assures that the results are not driven by funds following a particular trading style. The portfolios are equally weighted and rebalanced annually.

Panel A of Table 7 presents the performance of the trading strategies of interest. For each portfolio the table reports the mean return in percentage per year, standard deviation, Sharpe ratio, information ratio, maximum portfolio drawdown, the Fung and Hsieh (2004) alpha, and the pooled portfolio alpha. The Fung and Hsieh (2004) and pooled alphas for each strategy are computed via time-series regressions of the portfolio's realized excess returns on the relevant factors with constant factor loadings.[17] The table also shows the performance of hedge portfolios taking a long position in the top decile portfolio and a short position in the bottom decile for each indicator. Although hypothetical in nature, the performance of these hedge portfolios provides an economic measure of the ability of the corresponding alpha measures to distinguish between superior and inferior funds.

To begin, consider the performance of the trading strategies in panel A that rely on the optimal five-model pool to assess fund performance. The top decile portfolio delivers an average excess return of 5.26% per year. By contrast, the top decile strategies relying on the Fung and Hsieh (2004), stepwise, and style-adjusted models earn only 4.52%, 3.76%, and 4.97% per year, respectively. The pooled top decile portfolio also exhibits better performance in terms of Sharpe ratio and information ratio. The pooled investment strategy achieves this superior performance while exposing investors to less risk. The annual standard deviation for the top decile portfolio based on pooled alpha is 5.01%, and the maximum drawdown is 15.9%. Both of these risk measures are lower in magnitude than the corresponding figures for the other top decile strategies. The pooling approach is also effective in identifying poor performers. The bottom decile of funds according to the pooled benchmark underperforms the corresponding portfolios for the other three indicators based on mean return, Sharpe ratio, information ratio, and drawdown.

The pooled alpha and seven-factor alpha estimates reported in panel A of Table 7 provide further confirmation of the value of model pooling in forecasting fund performance. For example, the long-short strategy based on the optimal pool earns a pooled alpha of 2.84% per year and a seven-factor alpha of 3.42% per year. Both of these figures are statistically significant at the 5% level (t-statistics of 2.35 and 2.64, respectively). In contrast, the differences in estimated alphas for the top and bottom decile portfolios based on the Fung and Hsieh (2004) model, stepwise

[17] The pooled alphas are specifically computed as follows. We first regress the portfolio excess returns on the factors for each model and define the monthly model alpha as the sum of the estimated intercept and residual term. We then compute monthly pooled alphas using a weighted average of the monthly model alphas, with model weights equal to the average constituent model weights estimated using the prior 24 months of individual fund conditional densities. The pooled alpha t-statistics reported in Table 7 are based on the time-series variability of these monthly pooled portfolio alphas over the full sample period.

RIGHTSLINK

Table 7 Portfolios Formed on Past Performance Measures

Panel A: Performance measures for portfolios formed on alpha *t*-statistics

	Optimal pool			Fung and Hsieh (2004) model			Stepwise model			Style-adjusted model		
Performance measure	$P1$	$P10$	$P10-P1$	$P1$	$P10$	$P10-P1$	$P1$	$P10$	$P10-P1$	$P1$	$P10$	$P10-P1$
Mean (%)	2.59	5.26	2.66	4.32	4.52	0.21	3.51	3.76	0.25	3.08	4.97	1.89
Standard deviation (%)	7.67	5.01	5.04	7.22	5.55	5.68	8.32	5.81	5.65	6.77	8.82	7.22
Sharpe ratio	0.34	1.05	0.53	0.60	0.82	0.04	0.42	0.65	0.04	0.46	0.56	0.26
Information ratio	0.24	1.40	0.77	0.77	0.86	0.01	0.39	0.59	0.13	0.49	0.27	−0.06
Maximum drawdown (%)	27.3	15.9		18.7	21.4		27.1	19.7		24.0	26.6	
Pooled alpha (%)	1.35	4.19	2.84	3.34	3.09	−0.25	2.15	2.27	0.11	1.86	2.50	0.64
	(1.17)	(5.21)	(2.35)	(2.79)	(3.11)	(−0.17)	(1.78)	(2.29)	(0.08)	(1.75)	(1.85)	(0.41)
Seven-factor alpha (%)	1.04	4.46	3.42	3.28	3.33	0.05	1.68	2.30	0.62	2.02	1.64	−0.38
	(0.82)	(4.79)	(2.64)	(2.63)	(2.94)	(0.04)	(1.33)	(2.02)	(0.44)	(1.51)	(0.81)	(−0.19)

Panel B: Comparison of portfolio alphas

		Fung and Hsieh (2004) model		Stepwise model		Style-adjusted model	
Category	Performance measure	$\Delta(P10)$	$\Delta(P10-P1)$	$\Delta(P10)$	$\Delta(P10-P1)$	$\Delta(P10)$	$\Delta(P10-P1)$
All funds	Pooled alpha (%)	1.10	3.09	1.92	2.72	1.69	2.20
		(2.14)	(3.51)	(3.79)	(3.16)	(1.56)	(1.53)
	Seven-factor alpha (%)	1.13	3.36	2.16	2.80	2.82	3.79
		(2.00)	(3.52)	(3.98)	(3.06)	(2.02)	(2.11)
Directional funds	Pooled alpha (%)	2.25	7.40	4.94	7.74	1.54	1.93
		(1.48)	(3.15)	(3.35)	(3.55)	(0.58)	(0.48)
Nondirectional funds	Pooled alpha (%)	−1.70	0.21	−0.13	−0.50	−0.09	0.82
		(−1.35)	(0.10)	(−0.10)	(−0.23)	(−0.05)	(0.33)
Semidirectional funds	Pooled alpha (%)	2.01	3.18	1.57	2.18	2.84	2.99
		(2.58)	(2.59)	(2.15)	(1.72)	(1.99)	(1.74)
Fund of funds	Pooled alpha (%)	0.10	0.66	0.70	1.39	0.54	1.32
		(0.19)	(0.72)	(2.05)	(1.95)	(0.68)	(1.16)

Notes. The table reports mean excess returns, standard deviations, Sharpe ratios, information ratios, optimal pooled alphas, and Fung and Hsieh (2004) seven-factor model alphas in annualized terms, as well as drawdown percentages for portfolios formed on prior *t*-statistics from the optimal five-model pool, the Fung and Hsieh (2004) seven-factor model, the stepwise regression model, and the style-adjusted model. The *t*-statistics used to construct the trading strategies are computed using a rolling 24-month window of abnormal return data. Each January funds are sorted within each fund category (i.e., directional, nondirectional, and semidirectional) into decile portfolios based on each performance measure. Panel A reports the performance of trading strategies that invest in the bottom ($P1$) and/or top ($P10$) decile groups. Panel B compares the optimal five-model pool to each of the other three approaches in predicting hedge fund performance. This panel shows the difference in alphas for the top decile (i.e., $P10$) and hedge (i.e., $P10-P1$) portfolios for the strategy based on the optimal pool relative to the strategy based on each of the three alternative models (i.e., the Fung and Hsieh 2004 seven-factor model, the stepwise regression model, and the style-adjusted model). Panel B also reports differences in portfolio alphas by fund category. The portfolios are equally weighted. The numbers in parentheses are *t*-statistics.

model, and style-adjusted model are much smaller in economic magnitude and statistically insignificant in all cases. Despite their widespread use in evaluating hedge fund performance, these models appear to be less capable of reliably distinguishing between superior and inferior funds.[18]

Figure 2 provides a graphical representation of the performance differences among the trading strategies. The figure shows the growth of a \$1 million investment in the top decile portfolios and bottom decile portfolios based on pooled alphas, seven-factor alphas, stepwise regression alphas, and style-adjusted returns. At the end of the 13-year investment period, an investor following the top decile strategy based on pooled alpha would have seen her initial investment grow to \$2.68 million. In comparison, investors following the seven-factor, stepwise regression, and style-adjusted methods to assess manager performance would have ended the sample period with only \$2.42 million, \$2.19 million, and \$2.49 million, respectively. One notable feature of this plot is the large losses experienced by the seven-factor and style-adjusted top decile strategies around the onset of the recent financial crisis. For example, the seven-factor and style-adjusted portfolios experienced declines of 19.1% and 20.5%, respectively, during the period from

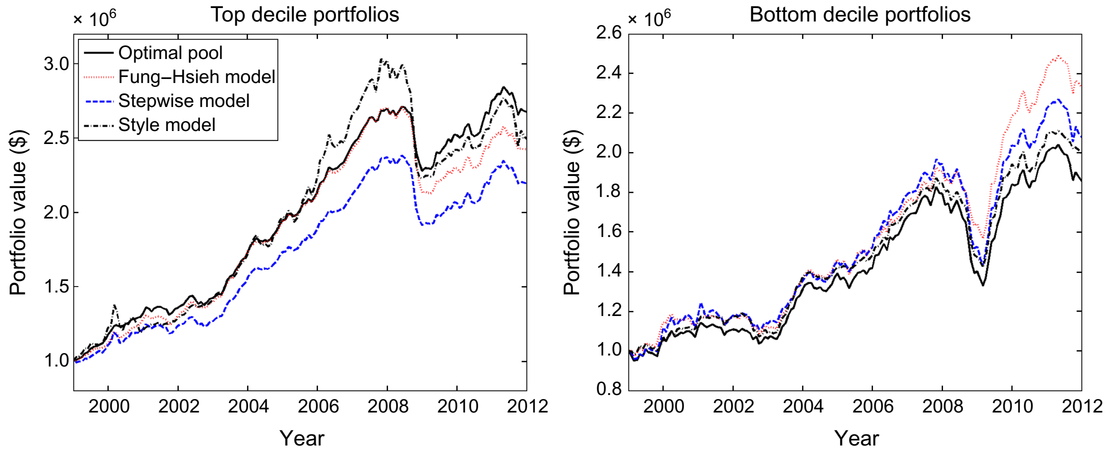
[18] In unreported results, we examined the ability of pooled alphas estimated over prior periods of either hedge fund market weakness or hedge fund market strength to distinguish between superior and inferior funds. We find that conditioning on broad hedge fund market performance does not qualitatively change the nature of the results in panel A of Table 7.

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Figure 2 (Color online) Growth of $1 Million Investment



Notes. The figure shows the growth of a $1 million initial investment in the top and bottom decile portfolios sorted on prior t-statistics from the optimal five-model pool, the Fung and Hsieh (2004) seven-factor model, the stepwise regression model, and the style-adjusted model. The portfolios are equally weighted and rebalanced at the beginning of January each year. The t-statistics used to construct the trading strategies are computed using a rolling 24-month window of abnormal return data for each fund.

September to December 2008. The pooled top decile portfolio experienced a more modest drop in value of only 14.4% over the same period. Finally, the bottom decile strategy according to the pooled indicator also results in a lower ending portfolio value ($1.86 million) when compared to the three alternatives.

Panel B of Table 7 provides additional tests on the usefulness of each modeling approach in forecasting future fund returns. This panel shows pooled and seven-factor alpha estimates for hypothetical portfolios taking a long position in the pooled top decile group and a short position in the top decile based on each of the three alternatives (i.e., $\Delta(P10)$). The table also reports the differences in performance between hedge portfolios formed on the various strategies (i.e., $\Delta(P10 - P1)$). Across all funds, the model pooling approach reliably outperforms the other three performance evaluation methods. The differences in top decile pooled alphas (seven-factor alphas) range from 110 to 192 (113 to 282) basis points per year. These estimates are statistically significant at the 5% level in five out of six cases. The reported differences in alphas for the hedge portfolios are also economically large, ranging from 220 to 379 basis points per year. These results further confirm that hedge fund evaluation based on model pooling leads to improved inference regarding fund performance.

Panel B of Table 7 also provides a breakdown of the investment strategy results by hedge fund category. The figures for funds following directional, semidirectional, and fund of funds strategies are qualitatively similar to the reported alphas for all funds in the top portion of the panel. In contrast, the value of following the model pooling approach appears to be absent among nondirectional funds. The seven-factor, stepwise, and style-adjusted strategies all deliver superior performance in terms of top decile alphas. These differences, however, are not statistically significant. One potential explanation for the lack of persistence for nondirectional funds with high pooled alphas could relate to the practical difficulty of scaling certain strategies such as convertible arbitrage and fixed income arbitrage (see, e.g., Getmansky et al. 2010).[19]

In summary, the results for the trading strategies presented in this section confirm the value of model pooling for hedge fund evaluation and investment decisions. More generally, these results serve to further highlight the problem of model error that plagues conventional performance evaluation methods that rely on a single benchmark model.

4.5. Failure Prediction

This subsection compares the ability of various performance measures in forecasting hedge fund failures. We are, thus, contrasting the ability of each of the modeling approaches to identify poor performing funds by focusing on failure rates rather than returns. As noted by Bollen (2013), to the extent that poorly performing funds may cease reporting their performance to a given hedge fund database, fund failure is a more straightforward indicator of poor performance.

Following the empirical approach in Liang and Park (2010) and Bollen (2013), we gauge the ability of each competing model or pool of models to predict failures using a Cox (1972) proportional hazards analysis with time-varying predictors. We identify a

[19] We are hesitant to read too much into these results, however, as compared to the other three style classifications, the nondirectional group contains by far the fewest number of funds. The resulting trading strategies for nondirectional funds summarized in panel B of Table 7 are, thus, relatively undiversified.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

hedge fund as failing if it is dropped from the sample with a drop reason of "fund liquidated" or "unable to contact fund" and has an average excess return over the prior 12 months that falls below the median average return for funds in the TASS database over the same period. Liang and Park (2010) demonstrate the importance of including a performance filter in characterizing hedge fund failures. Funds that leave the sample with alternative drop reasons from TASS or above-median excess returns over the previous year are treated as censored observations of live funds.

Identifying the exact timing of fund failure is not possible using the TASS database because some funds likely cease reporting their returns several months prior to failure. We therefore estimate our hazard model on a fund-year basis. We model the hazard function for a given hedge fund as

$$h(t, z) = h_0(t)\exp[\beta' z(t)]. \qquad (17)$$

The hazard rate gives the instantaneous rate of fund failure at time t conditional on survival up to time t. The baseline hazard rate, $h_0(t)$, depends only on the fund's age and is scaled up or down depending on the values of the time-varying covariates, $z(t)$. The primary question of interest in this analysis is the extent to which the various measures of past performance are informative in predicting fund failures. As such, for each of the four modeling approaches (i.e., pooling, seven-factor model, stepwise model, and style-adjusted returns), sample funds are ranked at the beginning of each year based on the alpha t-statistic computed from the prior two years of abnormal returns.[20] We then construct a *low t-statistic* indicator variable for each modeling approach that equals one for funds ranked in the lowest quartile. We define event-time zero as the date on which the first performance measure for a given fund becomes available. Our model therefore characterizes fund failures over the period from January 1999 to December 2011.

We rely on the prior literature on hedge fund failure prediction in choosing the remaining covariates. Following Liang and Park (2010), we include dummy variables to identify funds with a high-water mark provision, personal investment by fund managers, leverage, and lockup provisions. We also include the log of fund age in months as an explanatory variable. We control for recent past performance by including the percentile rank compared to all funds in existence based on average excess returns over the previous 12 months (return rank). To control for downside risk, we compute for each fund-year the expected shortfall, defined as the fund's mean monthly excess return conditional on the excess return falling below the fifth percentile using all prior observations (expected shortfall). Finally, we use dummy variables for each of the four broad investment styles (i.e., directional, nondirectional, semidirectional, and fund of funds) and also include annual fixed effects. All of the time-varying failure predictor variables are updated each year and are observable at the beginning of the relevant forecast period.

The results of the survival analysis are presented in panel A of Table 8. The first four models consider each of the individual *low t-statistic* indicator variables in turn while controlling for the other known predictors of hedge fund failure. The fifth model includes the four indicators of poor performance simultaneously to assess incremental effects. For each model covariate, the table provides the parameter estimate, p-value, and hazard ratio. In this context, the reported hazard ratios allow us to easily assess the economic significance of each covariate.

Results for the first model suggest that the *low t-statistic* dummy variable corresponding to the optimal pooled benchmark is a statistically significant predictor of hedge fund failure (p-value of 0.000). This coefficient estimate is also economically meaningful. The associated hazard ratio suggests that funds classified in the bottom quartile according to the pooled benchmark model have hazard rates that are 66% larger than those for other funds even after controlling for prior returns and downside risk. To provide some additional perspective on the magnitude of the marginal impact of this variable, note that the average annual failure rate for sample funds is 5.14%. If we set the baseline probability of failure at this rate, funds classified as poor performers according to the pooled benchmark experience a jump in expected rate of failure to 8.53% per year. The table also reports the p-value for the χ^2 likelihood ratio test comparing the log-likelihood values for the unrestricted model and a restricted model that omits the *low t-statistic* dummy variable. The restriction is rejected at the 1% level. For the remaining control variables, our results are generally in line with prior studies. Only prior return rank and personal investment by fund managers are significantly related to the probability of failure at the 5% level. As expected, better performing funds are more likely to survive. The positive coefficient on personal investment may reflect that managers following riskier strategies are required to invest more of their own capital as a signal to potential investors.

The three subsequent sets of results in panel A of Table 8 correspond to the failure predictors based on the seven-factor model, stepwise regression model, and style-adjusted model. The estimated coefficient

[20] For the pooled alpha t-statistics, the model weights are computed contemporaneously to the abnormal returns using the prior two years of predictive densities. This approach is identical to the one used to construct fund alpha t-statistics for the trading strategies in Table 7.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 8 Failure Prediction

Panel A: Predicting failure with performance measures

	(1)			(2)			(3)			(4)			(5)		
	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR
High-water mark	−0.124	0.147	0.88	−0.133	0.119	0.88	−0.127	0.137	0.88	−0.127	0.135	0.88	−0.125	0.144	0.88
Personal investment	0.205	0.016	1.23	0.215	0.012	1.24	0.212	0.013	1.24	0.216	0.012	1.24	0.207	0.016	1.23
Leverage	−0.031	0.713	0.97	−0.042	0.616	0.96	−0.030	0.722	0.97	−0.034	0.681	0.97	−0.027	0.742	0.97
Lockup provision	−0.132	0.190	0.88	−0.140	0.166	0.87	−0.133	0.188	0.88	−0.139	0.169	0.87	−0.122	0.223	0.89
Age	−0.340	0.436	0.71	−0.304	0.486	0.74	−0.337	0.440	0.71	−0.314	0.474	0.73	−0.372	0.394	0.69
Return rank	−2.521	0.000	0.08	−2.791	0.000	0.06	−2.669	0.000	0.07	−2.651	0.000	0.07	−2.525	0.000	0.08
Expected shortfall	−0.752	0.326	0.47	−0.217	0.775	0.81	−0.594	0.434	0.55	0.046	0.950	1.05	−0.624	0.417	0.54
Low t-statistic, pool	0.507	0.000	1.66										0.571	0.000	1.77
Low t-statistic, Fung–Hsieh				0.114	0.192	1.12							−0.413	0.000	0.66
Low t-statistic, stepwise							0.377	0.000	1.46				0.243	0.026	1.28
Low t-statistic, style adjusted										0.278	0.001	1.32	0.119	0.206	1.13
Likelihood ratio test	34.187	0.000		1.692	0.193		19.520	0.000		10.308	0.001		49.741	0.000	

Panel B: Predicting failure with model weights

	(6)			(7)			(8)			(9)			(10)		
	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR	Coeff.	*p*-value	HR
High-water mark	−0.118	0.169	0.89	−0.125	0.144	0.88	−0.125	0.144	0.88	−0.118	0.168	0.89	−0.119	0.163	0.89
Personal investment	0.214	0.013	1.24	0.204	0.017	1.23	0.204	0.017	1.23	0.210	0.014	1.23	0.215	0.012	1.24
Leverage	−0.035	0.672	0.97	−0.030	0.719	0.97	−0.032	0.699	0.97	−0.041	0.622	0.96	−0.033	0.695	0.97
Lockup provision	−0.129	0.201	0.88	−0.129	0.202	0.88	−0.137	0.176	0.87	−0.127	0.209	0.88	−0.132	0.190	0.88
Age	−0.342	0.434	0.71	−0.342	0.433	0.71	−0.344	0.430	0.71	−0.367	0.401	0.69	−0.328	0.453	0.72
Return rank	−2.514	0.000	0.08	−2.528	0.000	0.08	−2.502	0.000	0.08	−2.475	0.000	0.08	−2.528	0.000	0.08
Expected shortfall	−0.675	0.377	0.51	−0.773	0.314	0.46	−0.704	0.360	0.50	−0.614	0.424	0.54	−0.629	0.409	0.53
Low t-statistic, pool	0.514	0.000	1.67	0.507	0.000	1.66	0.503	0.000	1.65	0.519	0.000	1.68	0.500	0.000	1.65
Weight, domestic equity	−0.161	0.232	0.85												
Weight, FI and commodity				−0.110	0.435	0.90									
Weight, global							−0.135	0.296	0.87						
Weight, trend following										0.308	0.071	1.36			
Weight, index option													0.478	0.009	1.61
Likelihood ratio test	1.446	0.229		0.616	0.433		1.106	0.293		3.147	0.076		6.355	0.012	

Notes. The table reports estimation results for Cox (1972) proportional hazards models with time-dependent covariates. We estimate each model on a fund-year basis. For each variable, the table provides parameter estimates, *p*-values, and hazard ratios (HR). The models include style and year fixed effects. Failing funds are those that are dropped from the sample with a *DropReasonid* equal to 1 (fund liquidated) or 3 (unable to contact fund) and an average excess return over the prior 12 months that falls below the median average return for funds in the TASS database over the same period. *Low t-statistic* for a given indicator (i.e., pooled model, Fung and Hsieh 2004 model, stepwise model, or style-adjusted model) is a dummy variable equal to 1 for funds that rank in the lowest quartile based on the corresponding alpha *t*-statistic from the prior 24 months. *Weight* for a given model is the weight assigned to that model in the optimal five-model pool using the prior 24 months of conditional densities. For each of the 10 models, the table reports a *p*-value for the χ^2 likelihood ratio test for the difference in log-likelihood values of the unrestricted model and a restricted model that omits all of the *low t-statistic* dummy variables (panel A) or *weight* variables (panel B).

for the *low t-statistic* indicator variable for the seven-factor model is positive but statistically insignificant (*p*-value of 0.192). In contrast, the stepwise and style-adjusted indicators are strongly significant (*p*-values of 0.000 and 0.001, respectively). Compared to the pooled benchmark indicator variable (i.e., model (1)), however, the economic importance of these predictors is more limited. The hazard ratios imply marginal impacts on failure probabilities ranging from only 12% to 46%. The fifth model includes all four of the poor performance indicators. The pooled indicator variable retains its statistical and economic significance. In contrast, the dummy variable for the seven-factor model reverses sign, and the style-adjusted predictor becomes insignificant at the 5% level (*p*-value of 0.206) in forecasting hedge fund failures. The stepwise model predictor remains significant (*p*-value of 0.026), but its hazard ratio declines from 1.46 to 1.28 in the presence of the pooled performance measure. The estimated marginal effect for the pooled dummy variable remains economically large as indicated by the reported hazard ratio of 1.77. These results suggest that a low pooled alpha is a more reliable predictor of failure than a low alpha from the competing performance attribution models.

Finally, we also examine the extent to which the estimated optimal model weights have predictive power for failures among sample funds. For example,

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Table 3 shows that the trend following and index option models receive the smallest average weight across all funds. These models capture fund exposure to nonlinear factors and, as such, may become much more relevant during times of market stress. Panel B of Table 8 confirms that this is the case. The models reported in this panel include each of the five time-varying optimal model weights in turn. The weights for a given fund are estimated using the prior 24 months of conditional densities. The weights assigned to the domestic equity, fixed income and commodity, and global models (models (6) to (8)) are not significant predictors of fund failure. In contrast, the weights assigned to the trend following and index option models are both associated with higher hazard rates for sample funds (p-values of 0.071 and 0.009, respectively). Thus, the model pooling approach yields predictive content for failure likelihood in both the estimates of lagged performance and the model weights.

5. Concluding Remarks

The conventional approach of using a single benchmark model to evaluate hedge fund performance is problematic given the flexibility in hedge fund strategies. As a result, inference about skill is inevitably contaminated by the error in the benchmark model. In contrast to the conventional method, this paper proposes a model pooling approach: it employs a benchmark obtained by combining a set of diverse models to evaluate hedge fund performance. The weights assigned to the individual models in the pool are based on the log score criterion, a measure of the conditional performance of a model. Intuitively, the optimal pooled benchmark represents a combination of models that mitigates (benchmark) model error. We illustrate the advantages of a pooled benchmark, in both statistical and economic terms, over conventional benchmarks including the Fung and Hsieh (2004) model, stepwise regression methods, and style-adjusted approaches. In particular, we document the benefits of using a pooled benchmark in a real-time fund selection and investment strategy, hedge fund replication, and fund failure prediction.

Acknowledgments

The authors thank Vikas Agarwal, Nicolas Bollen (Western Finance Association discussant), John Geweke, Hao Jiang, Wei Jiang (the department editor), Bing Liang, an associate editor, and two anonymous referees, as well as seminar participants at Iowa State University, the University of Iowa, the 2013 Western Finance Association meetings, and the Sixth Conference on Professional Asset Management at Erasmus University for helpful comments and suggestions. The authors thank Alex Olsen for assistance with the Lipper TASS hedge fund database. Michael O'Doherty acknowledges financial support from the University of Missouri Research Board.

References

Agarwal V, Naik NY (2000a) Multi-period performance persistance analysis of hedge funds. *J. Financial Quant. Anal.* 35:327–342.

Agarwal V, Naik NY (2000b) On taking the alternative route: Risks, rewards, and performance persistence of hedge funds. *J. Alternative Investments* 2:6–23.

Agarwal V, Naik NY (2004) Risks and portfolio decisions involving hedge funds. *Rev. Financial Stud.* 17:63–98.

Amisano G, Giacomini R (2007) Comparing density forecasts via weighted likelihood ratio tests. *J. Bus. Econom. Statist.* 25: 177–190.

Ang A, Gorovyy S, van Inwegen GB (2011) Hedge fund leverage. *J. Financial Econom.* 102:102–126.

Avramov D, Kosowski R, Naik NY, Teo M (2011) Hedge funds, managerial skill, and macroeconomic variables. *J. Financial Econom.* 99:672–692.

Bali TG, Brown SJ, Caglayan MO (2012) Systematic risk and the cross section of hedge fund returns. *J. Financial Econom.* 106:114–131.

Baquero G, Ter Horst J, Verbeek M (2005) Survival, look-ahead bias and the persistence in hedge fund performance. *J. Financial Quant. Anal.* 40:493–517.

Bates JM, Granger CWJ (1969) The combination of forecasts. *Oper. Res. Quart.* 20:451–468.

Bollen NPB (2013) Zero-R^2 hedge funds and market neutrality. *J. Financial Quant. Anal.* 48:519–547.

Bollen NPB, Whaley RE (2009) Hedge fund risk dynamics: Implications for performance appraisal. *J. Finance* 64:985–1035.

Brown SJ, Goetzmann WN, Ibbotson RG (1999) Offshore hedge funds: Survival and performance 1989–1995. *J. Bus.* 72:91–118.

Capocci D, Hubner G (2004) Analysis of hedge fund performance. *J. Empirical Finance* 11:55–89.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *Internat. J. Forecasting* 5:559–581.

Cox DR (1972) Regression models and life-tables. *J. Royal Statist. Soc. Ser. B (Methodological)* 34:187–220.

Diebold FX, Lopez JA (1996) Forecast evaluation and combination. Maddala GS, Rao CR, eds. *Statistical Methods in Finance*, Vol. 14 (Elsevier Science B.V., Amsterdam), 241–268.

Fama EF, French KR (1993) Common risk factors in the returns on stocks and bonds. *J. Financial Econom.* 33:3–56.

Fama EF, French KR (2010) Luck versus skill in the cross-section of mutual fund returns. *J. Finance* 65:1915–1947.

Ferson WE (2010) Investment performance evaluation. *Ann. Rev. Financial Econom.* 2:207–234.

Ferson WE (2013) Ruminations on investment performance measurement. *Eur. Financial Management* 19:4–13.

Fung W, Hsieh DA (1997) Empirical characteristics of dynamic trading strategies: The case of hedge funds. *Rev. Financial Stud.* 10:275–302.

Fung W, Hsieh DA (2000) Performance characteristics of hedge funds and commodity funds: Natural versus spurious biases. *J. Financial Quant. Anal.* 35:291–307.

Fung W, Hsieh DA (2001) The risk in hedge fund strategies: Theory and evidence from trend followers. *Rev. Financial Stud.* 14: 313–341.

Fung W, Hsieh DA (2004) Hedge fund benchmarks: A risk-based approach. *Financial Analysts J.* 60:65–80.

Fung W, Hsieh DA, Naik NY, Ramadorai T (2008) Hedge funds: Performance, risk, and capital formation. *J. Finance* 63: 1777–1803.

Gelman A, Hwang J, Vehtari A (2014) Understanding predictive information criteria for Bayesian models. *Statist. Comput.* 24(6):997–1016.

RIGHTSLINK

Downloaded from informs.org by [155.246.103.35] on 05 April 2017, at 23:17 . For personal use only, all rights reserved.

Getmansky M, Liang B, Schwarz C, Wermers R (2010) Share restrictions and investor flows in the hedge fund industry, Working paper, University of Massachusetts, Amherst.

Getmansky M, Lo AW, Makarov I (2004) An econometric model of serial correlation and illiquidity in hedge fund returns. *J. Financial Econom.* 74:529–610.

Geweke J, Amisano G (2011) Optimal prediction pools. *J. Econometrics* 164:130–141.

Giacomini R, White H (2006) Tests of conditional predictive ability. *Econometrica* 74:1545–1578.

Glode V, Green RC (2011) Information spillovers and performance persistence for hedge funds. *J. Financial Econom.* 101:1–17.

Goetzmann WN, Ingersoll JE, Ross SA (2003) High-water marks and hedge fund management contracts. *J. Finance* 58: 1685–1717.

Good IJ (1952) Rational decisions. *J. Royal Statist. Soc. Ser. B* 14: 107–114.

Guidolin M, Timmermann A (2009) Forecasts of U.S. short-term interest rates: A flexible forecast combination approach. *J. Econometrics* 150:297–311.

Hall SG, Mitchell J (2007) Combining density forecasts. *Internat. J. Forecasting* 23:1–13.

Jagannathan R, Malakhov A, Novikov D (2010) Do hot hands exist among hedge fund managers? An empirical evaluation. *J. Finance* 65:217–255.

Jose VRR, Nau RF, Winkler RL (2008) Scoring rules, generalized entropy, and utility maximization. *Oper. Res.* 56:1146–1157.

Klein RW, Brown SJ (1984) Model selection when there is "minimal" prior information. *Econometrica* 52:1291–1312.

Kosowski R, Naik NY, Teo M (2007) Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *J. Financial Econom.* 84: 229–264.

Lan Y, Wang N, Yang J (2013) The economics of hedge funds. *J. Financial Econom.* 110:300–323.

Liang B (1999) On the performance of hedge funds. *Financial Analysts J.* 55:72–85.

Liang B, Park H (2010) Predicting hedge fund failure: A comparison of risk measures. *J. Financial Quant. Anal.* 45:199–222.

Newbold P, Harvey DI (2002) Forecast combination and encompassing. Clements MP, Hendry DF, eds. *A Companion to Economic Forecasting* (Blackwell, Malden, MA), 268–283.

O'Doherty M, Savin NE, Tiwari A (2012) Modeling the cross section of stock returns: A model pooling approach. *J. Financial Quant. Anal.* 47:1331–1360.

Pastor L, Stambaugh RF (2003) Liquidity risk and expected stock returns. *J. Political Econom.* 111:642–685.

Patton A, Ramadorai T (2013) On the high-frequency dynamics of hedge fund risk exposures. *J. Finance* 68:597–635.

Rapach DE, Strauss JK, Zhou G (2010) Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Rev. Financial Stud.* 23:821–862.

Sharpe WF (1992) Asset allocation: Management style and performance measurement. *J. Portfolio Management* 18:7–19.

Stock JH, Watson MW (2003) Forecating output and inflation: The role of asset prices. *J. Econom. Literature* 41:788–829.

Stock JH, Watson MW (2004) Combination forecasts of output growth in a seven-country data set. *J. Forecasting* 23: 405–430.

Sun Z, Wang A, Zheng L (2012) The road less traveled: Strategy distinctiveness and hedge fund performance. *Rev. Financial Stud.* 25:96–143.

Timmermann A (2006) Forecast combination. Elliott G, Granger CWJ, Timmermann A, eds. *Handbook of Economic Forecasting* (North–Holland, Amsterdam), 135–196.

Titman S, Tiu C (2011) Do the best hedge funds hedge? *Rev. Financial Stud.* 24:123–168.

Zellner A (1971) *An Introduction to Bayesian Inference in Econometrics* (John Wiley & Sons, Hoboken, NJ).

RIGHTSLINK