Original Article

Chasing performance persistence of hedge funds – Comparative analysis of evaluation techniques

Received (in revised form): 5th November 2008

Eero J. Pätäri
is Professor of Finance at the School of Business at Lappeenranta University of Technology (Finland) and a Partner of Confido Capital. He has published a large number of papers in journals such as the *Journal of Risk,* the *International Research Journal of Finance and Economics,* the *Journal of Performance Measurement,* and the *Journal of Money, Investment and Banking.* His research interests include performance of equity investment strategies, measurement techniques of portfolio risk and performance measurement, particularly performance persistence of mutual funds and hedge funds. In addition to his academic duties, he has consulted for management of several institutional investors in the field of risk and performance measurement.

Jussi Tolvanen
is Investment Specialist for a large Scandinavian financial institution based in Sweden.

Correspondence: Eero J. Pätäri Lappeenranta School of Business, Lappeenranta University of Technology, PO Box 20, FI-53851 LAPPEENRANTA, Finland
E-mail: Eero.Patari@lut.fi

ABSTRACT **We apply two complement methodologies (that is, stacked cross-sectional regression and quartile portfolio approach) in detecting the performance persistence of five different hedge fund styles. In addition, we compare the results obtained by using model-free performance metrics to those obtained by using both standard alphas of multifactor models and their empirical Bayesian counterparts. The results show that both the degree and existence of performance persistence vary among hedge fund styles and, in addition, depend on performance metric employed. Based on the combination of 3-year selection period and the subsequent 1-year holding period, model-free performance metrics (such as the Sharpe ratio and its downside risk-based variants) are more sensitive to detecting performance persistence than are factor-based performance metrics. Correspondingly, the prediction power of empirical Bayesian alphas is better than that of standard OLS alphas. The strongest evidence of performance persistence within the sample is among event-driven funds, for which 10 out of 12 persistence tests performed indicate significant results.**

Journal of Derivatives & Hedge Funds (2009) **15,** 223–240. doi:10.1057/jdhf.2009.11

Keywords: hedge funds; performance persistence; Bayesian estimation; Bayesian alphas; performance measures; factor models

INTRODUCTION

Along with the rapid growth of the hedge fund industry and the emergence of hedge fund databases, scholars have started to study the persistence of hedge fund performance, analogously to the trend experienced in the mutual fund literature during the past 40 years. As can be judged from the abundant mutual fund literature over several decades, detecting performance persistence is challenging in mutual fund samples as a result of the many potential sources of bias (for a summary of potential biases in the context of mutual fund performance persistence studies, see Pätäri[1]). However, evaluating the persistence of hedge fund performance is even more challenging, as the diversity of hedge funds is even greater than it is among mutual funds. For example, hedge fund returns do not follow any conventional distributions, given the funds' dynamic trading strategies and their holdings of derivative-type securities or agreements. In addition, the complexity of hedge fund strategies increases the risk of spurious persistence findings caused by model misspecification. Moreover, the higher attrition rate of hedge funds strengthens many of the biases documented in the mutual fund literature.

The evidence of performance persistence among hedge funds is mixed. Using annual return data and a sample of only offshore funds, Brown *et al*[2] find hardly any evidence of persistence after controlling for style effect, while the results of Agarwal and Naik[3] show short-term persistence at the quarterly horizon in the traditional two-period framework but no persistence at the annual horizon in the multi-period framework. Harri and Brorsen[4] find some evidence of persistence whose strength varies among hedge fund styles. The results show the greatest persistence for market neutral and for two funds of fund styles, and weaker persistence for the event driven, global and global macro styles. However, for some other styles no evidence of persistence is found. In this sense, the results of Harri and Brorsen are consistent with those of Agarwal and Naik,[5] who also document that some hedge fund styles exhibit stronger performance persistence than others.

Controlling for look-ahead bias, Baquero *et al*[6] find a clear pattern of persistence in raw returns of hedge funds at the quarterly horizon but no statistically significant persistence at the annual horizon. Based on style-adjusted returns, the results show (on average) outperformance for top-decile funds of the preceding selection period at both the quarterly and annual horizons, though not significantly in the statistical sense. In addition, the authors note that persistence in hedge fund performance seems to be located in both the top and bottom parts of distribution. In this sense, the results of Baquero *et al* are in contrast with those of Capocci *et al*,[7] who find that persistence, if any, is mostly located among medium performers. Gregoriou *et al*[8] show that fund portfolios made up of non-directional funds with the highest Information ratio and/or Sharpe ratio are likely to exhibit a remarkable persistence and continue to dominate the best fund of funds on the basis of the three performance metrics being examined (that is, the Sharpe ratio, 9-factor alpha and the corresponding Information ratio).

Kosowski *et al*[9] are the first to evaluate the persistence of hedge fund performance using Bayesian alphas. The Bayesian approach in the performance measurement context was first introduced in the mutual fund literature by Baks *et al*[10] and Pástor and Stambaugh,[11] and

Kosowski *et al* applied the Bayesian approach of Pástor and Stambaugh, which takes advantage of prior information in seemingly unrelated assets to overcome short sample problems, and to improve the precision of performance estimates. Their results show that hedge fund performance persists at annual horizons, and, further, that the prediction power of past alphas can be significantly improved by employing Bayesian alphas in forming the decile portfolios.

The parallel results of the improved predictive accuracy stemming from the use of Bayesian methods are also reported in many studies that examine persistence in mutual fund performance (for example, see Bollen and Busse,[12] Busse and Irvine,[13] Huij and Verbeek[14]). Many variants of the Bayesian estimation methods have been employed in the previous studies. This article contributes the existing literature in applying the so-called empirical Bayesian approach for the first time in the hedge fund context. Most of the Bayesian approaches require prior information related to such issues as funds' expenses, investors' beliefs about managerial skill, or benchmark pricing abilities, or the returns on other mutual funds, benchmark factors or non-benchmark factors, while the empirical Bayesian method is based entirely on a cross-section of fund returns within the sample period. The basic principle is to use cross-sectional data as prior information, in which case the resulting belief in managerial skill is no longer fully subjective, but instead, entirely data based. For example, Jones and Shanken,[15] and Busse and Irvine[13] demonstrate that the predictive accuracy of Bayesian alphas is greatly affected by the prior beliefs that are attributed to the investor. The empirical Bayesian approach was applied first in the mutual fund context by Huij and Verbeek,[14] who document the superior prediction power of Bayesian alphas over standard OLS alphas.

We compare the prediction power of three performance metrics (that is, a standard frequentist style-adjusted 9-factor alpha, a corresponding Bayesian alpha and the Sharpe Ratio) to determine whether the performance persistence of hedge funds is dependent on the performance metrics employed. In addition, we use two complement methodologies in detecting the performance persistence of five different hedge fund styles (that is, convertible arbitrage, commodity trading advisor (CTA, henceforth), equity market neutral, event driven and global macro). First, we test the prediction power of the selection period performance on the holding period performance by means of cross-sectional regression. Second, we form top- and bottom-quartile portfolios based on selection period performance and test their holding period performance difference.

METHODOLOGY

Several models have been suggested to evaluate the performance of hedge funds. Typically, the multifactor models that are aimed to capture the common risk factors of diversified portfolios of hedge funds are employed for this purpose. The most widely used model of this type is the Fung-Hsieh 7-factor model, which includes the following factors: S&P 500 return in excess of the risk-free rate, Wilshire small cap minus large cap return, month-end to month-end change in the US Federal reserve 10-year constant maturity yield, corresponding change in the difference between Moody's Baa yield and the Fed's 10-year constant maturity yield, bond PTFS, currency PTFS, and commodity PTFS, where PTFS denotes primitive trend following strategy. While these seven factors can explain up to 80 per cent of monthly return variations for diversified hedge fund portfolios (see Fung and Hsieh[16]), their explanatory power with regard to returns of individual hedge funds is much lower. Therefore, we add two additional factors to the Fung-Hsieh model to improve its feasibility for hedge fund portfolio selection; first, as many hedge funds follow global stock investment strategies, we add a factor based on returns of Morgan Stanley Capital International world equities excluding US equities.[17] Second, as we examine each hedge fund class separately, the style factor for each fund category is included in order to extract the funds' true abnormal performance, that is, the performance beyond following a certain style. We use five different style factors for five corresponding hedge fund styles.

We employ three performance metrics in our comparative analysis: a standard frequentist style-adjusted 9-factor alpha, a corresponding Bayesian alpha and the Sharpe Ratio. At the outset of the study, the Fung-Hsieh 7-factor alpha was also employed, but was excluded from the comparative analysis, as the average adjusted coefficients of determination were remarkably higher for the 9-factor regression models including two additional factors than they were for the Fung-Hsieh 7-factor models.[18] At the outset, we also employed several performance metrics that take account of skewness and kurtosis of fund return distributions. This was done because hedge funds frequently generate non-normal return distributions, and therefore the applicability of the Sharpe ratio to the evaluation of hedge funds particularly as the sole or the primary performance metric is questioned in many previous studies (for example, see Fung and Hsieh,[19] Lo[20] 2002, Brooks and Kat,[21] Gregoriou and Gueyie,[22] Mahdavi,[23] Sharma,[24] Morton *et al*[25]). However, the recent findings of Eling and Schuhmacher[26] show that despite significant deviations of hedge fund returns from a normal distribution, the Sharpe ratio results in rank orders that are virtually identical to 12 other performance measures that are based on downside risk. In addition, the evidence from mutual fund markets indicates that performance rank orders are not very sensitive to the selection of risk measure (for example, see Ferruz *et al*[27] for a case of money market funds, and Ferruz *et al*[28] and Pätäri[29] for a case of equity funds). We compared the rank orders produced by the Sharpe ratio with rankings based on one variant of the Sortino ratio, a modified Sharpe ratio that uses modified Value at Risk (VaR) as a risk metric (for details, see Favre and Signer[30]), and an adjusted Sharpe ratio that is closely related to a modified Sharpe ratio but that, in addition, can capture the occasional negative risk premiums and positive modified VaR figures that sometimes occurred in the sample employed. After numerous experiments, we came to the same conclusion as Eling and Schuhmacher[26]: for the sample employed, the hedge fund portfolios formed on the basis of performance metrics capturing downside risk were very similar to those based on the Sharpe ratio. Therefore, these approaches were not included in the comparative analysis, as they would have added hardly any value into it. In contrast, the results from the quartile portfolio approach based on the Sharpe ratio can, under these conditions, be generalized to also hold also for the above-mentioned performance metrics that put more emphasis on downside risk.

The purpose of the comparative analysis is to determine whether the investor would benefit from using more sophisticated methods while selecting a hedge fund portfolio based on past performance. We employ two different methodologies. First, we run a cross-sectional regression throughout the full sample period to determine whether the performance figures from the selection period explain those from the holding period. Second, to get a more in-depth view of the characteristics of the potential consistency in hedge fund performance, we form top-quartile portfolios and bottom-quartile portfolios based on selection period performance and compare their performance in the subsequent holding period in order to examine whether the performance difference between past outperformers and past underperformers remains. Taking account of implications of rebalancing, continuous, stacked time-series of monthly returns for quartile portfolios are generated throughout the sample period. In both methodologies, the selection period performance is based on time-series of 36 monthly returns and the holding period performance on subsequent 12-month period time-series of returns.

As a representative total risk-based performance metric, we employ the traditional Sharpe ratio as follows:

$$\text{Sharpe Ratio} = \frac{R_i - R_f}{\sigma_i} \qquad (1)$$

where R_i = the average monthly return of a portfolio i; R_f = the average monthly risk free rate of the return; σ_i = standard deviation of the monthly excess returns of a portfolio i.

The 9-factor alpha represents the performance metrics that are based on the standard frequentist multi-factor model, and indicates the abnormal performance of the hedge fund/hedge fund portfolio i after controlling for its systematic risk exposures. The regression equation for calculating alpha is as follows:

$$r_t^i = \alpha^i + \sum_{k=1}^{K} \beta_k^i F_{k,t} + \varepsilon_t^i \qquad (2)$$

where r_t^i is the fund i excess return at time t, α^i is the intercept, β_k^i is the factor loading of hedge fund i on factor k, $F_{k,t}$ is the return of factor k at time t and ε_t^i is the error term.

In order to calculate the corresponding empirical Bayesian alphas, we first specify the cross-sectional distribution of alphas and betas as normal:

$$\boldsymbol{\theta}_i \sim N(\boldsymbol{\mu}, \Sigma) \qquad (3)$$

where $\boldsymbol{\theta}_i = (\alpha_i, \beta_{1i}, \ldots, \beta_{ki})'$ $\boldsymbol{\mu}$ is the $(k+1)$ vector of the cross-sectional means of each estimate included in $\boldsymbol{\theta}_i$, and $\boldsymbol{\Sigma}$ denotes a $(k+1)$ by $(k+1)$ covariance matrix of the previous OLS estimates. Under the assumption of independent and identically distributed (i.i.d.) normal error terms received using OLS regression, the posterior distribution of θ_i is normal, with expected value given by

$$\boldsymbol{\theta}_i^* = \left(\frac{1}{\boldsymbol{\sigma}_i^2}\mathbf{X}_i'\mathbf{X}_i + \sum\nolimits^{-1}\right)^{-1}\left(\frac{1}{\boldsymbol{\sigma}_i^2}\mathbf{X}_i'\mathbf{X}_i\hat{\theta}_i + \sum\nolimits^{-1}\boldsymbol{\mu}\right) \qquad (4)$$

where $\hat{\theta}_i$ denotes the OLS estimate, σ_i^2 variance of the error terms, $\mathbf{X}_i$ the excess return matrix of the benchmark factors including the intercept, and $((\mathbf{1}/\boldsymbol{\sigma}_i^2)\mathbf{X}_i'\mathbf{X}_i + \mathbf{\Sigma}^{-1})^{-1}$ denotes the corresponding covariance matrix.

The estimation of θ_i^* involves parameters (that is, μ, σ_i and $\sum$) that cannot be estimated without knowing θ_i^*. This creates a recursive problem that Hu and Maddala[31] overcome by using an empirical Bayesian approach. They calculate the parameters μ, σ_i^2 and Σ iteratively using $\hat{\boldsymbol{\theta}}_i$ as initial estimation of $\boldsymbol{\theta}_i^*$:

$$\boldsymbol{\mu}^* = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\theta}_i^* \quad (5)$$

$$\boldsymbol{\sigma}_i^{2*} = \frac{1}{T_i - k}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_i^*)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_i^*) \quad (6)$$

$$\boldsymbol{\Sigma}^* = \frac{1}{N-1}\left[\boldsymbol{D} + \sum_{i=1}^{N}(\boldsymbol{\theta}_i^* - \boldsymbol{\mu})(\boldsymbol{\theta}_i^* - \boldsymbol{\mu})'\right] \quad (7)$$

where N is the number of funds, T_i is the number of observations of fund i, $\mathbf{y}_i$ is a vector containing the excess returns of fund i, and $\mathbf{D}$ denotes a diagonal matrix with small positive entries (for example, entries of 0.00001 can be used to improve the convergence in the iterative procedure). After calculating the parameters in (5), (6) and (7), $\boldsymbol{\theta}_i^*$ is estimated again until the desired level of convergence of the parameters is obtained. Consistent with the previous studies employing the Bayesian estimation techniques, the Bayesian alphas are estimated only for selection periods, while the performance of holding period is determined on the basis of the Sharpe ratio and a standard frequentist 9-factor alpha only.

Statistical tests and adjustments

The statistical significances of differences between comparable pairs of the Sharpe ratios are given by P-values of the Ledoit-Wolf test,[32] which is based on the circular block bootstrap method. We also test the statistical significance of differences between portfolio alphas by applying the alpha spread test as follows:

$$t = \frac{\alpha_i - \alpha_i}{\sqrt{SE_{\alpha i}^2 + SE_{\alpha j}^2}} \quad (8)$$

where α_* is the alpha of portfolio*; $SE_{\alpha *}$ is the standard error of portfolio*.

The degrees of freedom for the test statistic are given as

$$= \frac{(SE_{\alpha i}^2 + SE_{\alpha j}^2)^2}{\frac{SE_{\alpha i}^4}{v_i} + \frac{SE_{\alpha j}^4}{v_j}} \quad (9)$$

where v_i and v_j are the degrees of freedom determined on the basis of the number of time-series returns in samples i and j ($v = n-1$).

Throughout the study, we use Newey-West standard errors in statistical tests to avoid problems related to autocorrelation and heteroscedasticity. In addition, we performed the normality test by Jarque and Bera for regression residuals, but the assumption of their normality is not generally violated except for few random cases.

DATA

The sample data consist of monthly returns of 220 hedge funds (44 funds per each style) for the 1997–2006 period. In order to get as unbiased a sample as possible, the funds are selected randomly from the TASS database of those funds that have at least a 5-year return history. The minimum length requirement for the track record is set to 5 years, as we employ a 3-year selection period and follow the typical method used in hedge fund studies to eliminate backfill bias, which requires removing the first 12 monthly returns of every fund. To maintain better comparability between the two methodologies employed, we also require the full time-series of holding period returns from our sample funds, knowing that this raises some survivorship bias. However, because this kind of bias cannot be avoided in the cross-sectional sample by any means, we consider it a smaller trade-off than that stemming from the alternative consequence that the degree of survivorship bias would have varied in different persistence tests. On the other hand, as the attrition rate is highest among younger hedge funds (for example, see Brooks and Kat,[21] Brown *et al*,[33] Gregoriou,[34] Amin and Kat[35]), it would not have made much difference if we had required only a 4-year return history instead.

The S&P 500 Composite Index, the Wilshire Small Cap Index, the Wilshire Large Cap Index, US Treasury securities at 10-year constant maturity and Moody's Baa Bond Index are from Datastream. The MSCI World ex US (Wexus) index is from the MSCI database and the 3-month US Treasury Bill is from the Federal Reserve database. Indices for PTFSs are from the web pages of the Fuqua School of Business.[36] The hedge fund style indices are from the TASS database.

The first 3 years of data (that is, 1997–1999) are used only in evaluating selection period performance. Correspondingly, the stacked time-series of holding period returns for quartile portfolios ranges from the beginning of 2000 to the end of 2006, and includes 84 monthly observations throughout the 7-year period.

Table 1 reports the descriptive statistics of an average hedge fund within each fund style for a 36-month selection period and a 12-month holding period. Although there are great individual extreme values, the average values of descriptive statistics are surprisingly moderate. It is also noteworthy that skewness and kurtosis of return distributions are on average lower for shorter holding periods than for longer selection periods. Generally, the relatively low degree of average skewness and kurtosis may explain why the quartile portfolios formed on the basis of downside-risk variants of the Sharpe ratio did not deviate from those based on the standard Sharpe ratio. The rank order of hedge fund styles on the basis of average volatility of returns is the same for both periods; the returns of CTA-style funds are clearly the most volatile (their average annual volatility is 19.7 per cent in the selection period, and 18.6 per cent in the holding period), whereas those of equity market neutral-style are the least volatile. In addition, the lowest average minimum return and the highest maximum return are attributed to CTA style, whereas equity market neutral style gets the highest minimum and the lowest maximum return for both the selection and holding periods (on average).

THE RESULTS

Cross-sectional regression results

The strongest evidence of performance persistence is found within every hedge fund style when Sharpe ratios are used as input variables of regression analyses (Table 2). The results are highly significant for four hedge fund styles (that is, for convertible arbitrage, event driven, global macro and equity market neutral style at the 1 per cent level) and significant at the 10 per cent level for CTA style. However, the results are significant only for one style, that is, for equity market neutral at the 10 per cent level, when the regression is run with OLS alphas as input variables. When the Bayesian approach is applied to selection period alphas, the greatest difference in contrast

Table 1: Descriptive statistics of the hedge funds for (a) the 3-year selection period (January 1997 – December 2005) and (b) the 1-year holding period (January 2000 – December 2006)

	N	*Min.*	*Max.*	*Mean*	*SD*	*Skewness*	*Kurtosis*
Panel A							
Selection period							
Convertible Arbitrage	266	−0.047	0.058	0.008	0.021	−0.033	2.078
CTA	278	−0.119	0.142	0.008	0.057	0.175	0.730
Event Driven	277	−0.061	0.063	0.006	0.024	−0.316	2.962
Global Macro	258	−0.078	0.103	0.008	0.037	0.215	1.926
Equity Market Neutral	260	−0.037	0.049	0.006	0.019	0.100	1.002
Panel B							
Holding period							
Convertible Arbitrage	266	−0.024	0.038	0.007	0.017	0.012	0.641
CTA	278	−0.087	0.103	0.006	0.054	0.129	0.272
Event Driven	277	−0.030	0.038	0.005	0.019	−0.002	0.977
Global Macro	258	−0.049	0.067	0.007	0.033	0.097	0.601
Equity Market Neutral	260	−0.021	0.031	0.005	0.015	0.123	0.403

This table reports the descriptive statistics of the hedge fund sample taken from the TASS database. The sample includes 220 hedge funds in total (44 within each of the five styles). Panel A presents descriptive statistics for the 3-year selection period and Panel B those of the 1-year holding period. The statistics are based on monthly returns of hedge funds.

to the regression that is based solely on cross-sections of OLS alphas is observed for CTA style; employing the Bayesian alphas turns the non-persistence finding based on OLS alphas into performance reversal, which is significant at the 10 per cent level. In contrast, regression results based on the Sharpe ratios indicate almost equally significant persistence for the same style. In conclusion, the cross-sectional results for CTA style are paradoxical, and should be interpreted with extreme caution.

For the sample employed, the overall results of cross-sectional tests show that the Sharpe ratio is more sensitive detect performance persistence than are 9-factor alphas. For the pooled sample, 27.9 per cent of the variance of the holding period Sharpe ratio can be explained by its counterpart from the selection period, whereas for the other two factor-based performance metrics the explanatory power of past performance is virtually zero. For the fund style with the strongest persistence on the basis of cross-sectional results (that is, for equity market neutral), the explanatory power of past Sharpe ratios is double that of the average. However, it should be noted that except for convertible arbitrage and equity market neutral styles, the other

Table 2: Cross-sectional regression results

	N	*Slope*	*P-value*	R^2	*N*	*Slope*	*P-value*	R^2
		Sharpe ratio vs Sharpe ratio				*Sharpe ratio vs OLS alpha*		
Convertible Arbitrage	266	1.959	0.000★★★	0.314	266	0.005	0.025★★	0.020
CTA	278	0.325	0.088★	0.011	278	−0.014	0.523	0.002
Event Driven	277	0.391	0.000★★★	0.051	277	0.011	0.071★	0.013
Global Macro	258	0.401	0.001★★★	0.045	258	−0.006	0.537	0.002
Equity Market Neutral	260	0.932	0.000★★★	0.556	260	0.002	0.342	0.004
All styles	1339	1.400	0.000★★★	0.279	1 339	0.002	0.330	0.028
		OLS alpha vs OLS alpha				*OLS alpha vs Sharpe ratio*		
Convertible Arbitrage	266	0.114	0.282	0.005	266	8.801	0.420	0.003
CTA	278	−0.009	0.947	0.000	278	−0.616	0.636	0.001
Event Driven	277	−0.077	0.628	0.001	277	5.996	0.034★★	0.017
Global Macro	258	−0.043	0.740	0.000	258	−1.378	0.344	0.003
Equity Market Neutral	260	0.257	0.068★	0.014	260	0.306	0.951	0.000
All styles	1339	0.033	0.583	0.000	1 339	0.143	0.950	0.000
		Bayesian alpha vs OLS alpha				*Bayesian alpha vs Sharpe ratio*		
Convertible Arbitrage	266	0.219	0.421	0.003	266	58.163	0.037★★	0.017
CTA	278	−0.907	0.083★	0.011	278	9.693	0.043★★	0.016
Event Driven	277	−0.182	0.609	0.001	277	19.873	0.002★★★	0.038
Global Macro	258	0.468	0.274	0.005	258	6.378	0.187	0.006
Equity Market Neutral	260	0.461	0.042★★	0.018	260	3.556	0.657	0.001
All styles	1339	−0.131	0.655	0.000	1 339	18.002	0.005★★★	0.006

At the beginning of each year, realized performance measures over the subsequent 12 months (that is, holding period) are regressed on the performance measures over the preceding 36 months (that is, selection period). Performance is determined using 9-factor alphas and the Sharpe ratios. The holding period performance is estimated using the Sharpe ratios and standard OLS alphas, and the selection period performance using the Sharpe ratios, OLS alphas and Bayesian alphas. The analysis is conducted both on different types of funds separately and on the full sample of funds. The table reports the regression slope coefficients and their *P*-values (★★★, ★★ and ★ denote significance at the 1 per cent, 5 per cent and 10 per cent level, respectively), and R^2-values. The header line indicates the combination of explanatory and response variables used in each regression (for example, the header *Bayesian alpha versus Sharpe ratio* at the bottom right indicates that holding period Sharpe ratios are regressed on selection period Bayesian alphas).

coefficients of determination in style-specific cross-sectional regressions are also very low when cross-sectional tests are based on the Sharpe ratio.

In order to avoid spurious results stemming from the model misspecification bias,[37] we also test the power of model-free performance metric (represented in this context by the Sharpe ratio) to predict the holding period performance evaluated by measures based on factor models and vice versa. The results of these tests indicate that for some fund styles, the cross-predictions are better than those based on the same performance metrics for both periods (Table 2). For example, for event-driven style, both standard and Bayesian alphas predict the holding period Sharpe ratio better than they predict the holding period alpha. In addition, for convertible arbitrage and CTA styles, Bayesian alphas predict the holding period Sharpe ratios better than they predict OLS alphas (and also significantly in the statistical sense). Furthermore, the pooled sample results show that the dependence of the holding period Sharpe ratio on the selection period Bayesian alphas is significant at the 1 per cent level. However, for equity market neutral style, the cross-prediction power between model-free and factor-based models is distinctly weaker than the corresponding prediction power when the same performance metrics are used for both periods. Again, it should be noted that the highest coefficient of determination in cross-prediction tests is only 3.8 per cent, indicating the relatively low explanatory power of these tests in general.

Quartile portfolio approach

When the performance difference of top- and bottom-quartile portfolios is used as a persistence criterion, the strongest evidence of performance persistence is found within event-driven style, for which every variant of the top and bottom portfolio performance comparison employed shows highly significant results (Table 3). However, none of the variants employed detect persistence within global macro style.[38] For the three other hedge fund styles being examined, the results are somewhat mixed; for example, for CTA style, three out of six results are statistically significant and two other test statistics are also close to the 10 per cent significance level. For convertible arbitrage style, two significant results for persistence are documented when Bayesian alphas and the Sharpe ratio are employed as portfolio selection criteria, and the holding period performance is evaluated by the latter performance metric.

For equity market neutral style only one out of six statistical tests indicates significant persistence. The only significant result for this style is obtained when the Sharpe ratio is used as a performance metric for the selection and holding periods. These results deviate from those obtained from cross-sectional regression tests that indicate the strongest persistence particularly for this fund style. The explanation as to why persistence is not observed using the quartile portfolio approach when alphas (either OLS or Bayesian) are used as performance metrics is given by the performance reversal effect in the bottom-quartile portfolio of equity market neutral style. The holding period alphas of bottom portfolios are positive in both cases, and in a case when portfolios are formed on the basis of OLS alphas, the bottom-portfolio alpha is even higher than the top-portfolio alpha, though not significantly. However, the reversal effect among bottom portfolio funds is not strong enough to mitigate persistence in

Table 3: Performance comparison of top- and bottom-quartile style portfolios

Style portfolios	*Annual excess return (%)*	*Annual volatility (%)*	*Sharpe ratio*	$\tilde{d}^{\star,m}$*(sign.)*	*Alpha*	*Alpha spread (sign.)*	R^2
Panel A							
Convertible Arbitrage Q1	6.52	16.50	0.111	0.62	2.76%	3.61%	0.568
				(0.538)	(0.062★)	(0.224)	(0.000★★★)
Convertible Arbitrage Q4	2.10	11.92	0.051		−0.85%		0.514
					(0.741)		(0.000★★★)
CTA Q1	9.07	5.05	0.500	2.68	4.77%	6.57%	0.795
				(0.007★★★)	(0.127)	(0.134)	(0.000★★★)
CTA Q4	5.49	8.50	0.185		−1.80%		0.614
					(0.560)		(0.000★★★)
Event Driven Q1	5.49	4.42	0.350	2.82	2.11%	6.72%	0.624
				(0.005★★★)	(0.120)	(0.003★★★)	(0.000★★★)
Event Driven Q4	1.82	6.38	0.082		−4.61%		0.686
					(0.011★★)		(0.000★★★)
Global Macro Q1	5.46	7.01	0.221	0.79	−0.08%	−4.64%	0.470
				(0.428)	(0.973)	(0.189)	(0.000★★★)
Global Macro Q4	8.76	8.17	0.301		4.56%		0.512
					(0.085★)		(0.000★★★)
Equity Market Neutral Q1	3.85	4.52	0.243	0.87	0.98%	−0.83%	0.204
				(0.386)	(0.668)	(0.760)	(0.002★★★)
Equity Market Neutral Q4	3.77	2.72	0.393		1.81%		0.027
					(0.236)		(0.270)
Panel B							
Convertible Arbitrage Q1	8.84	4.07	0.608	2.96	4.54%	4.55%	0.455
				(0.003★★★)	(0.001★★★)	(0.112)	(0.000★★★)

Table 3 *continued*

Style portfolios	*Annual excess return (%)*	*Annual volatility (%)*	*Sharpe ratio*	$\tilde{d}^{\star m}$*(sign.)*	*Alpha*	*Alpha spread (sign.)*	R^2
Convertible Arbitrage Q4	6.88	8.68	0.226		−0.01%		0.549
					(0.998)		(0.000***)
CTA Q1	9.15	15.70	0.163	1.75	7.25%	10.91%	0.797
				(0.079*)	(0.015**)	(0.016**)	(0.000***)
CTA Q4	−1.42	12.73	−0.033		−3.67%		0.590
					(0.284)		(0.000***)
Event Driven Q1	7.01	3.52	0.557	5.15	4.12%	9.06%	0.580
				(0.000***)	(0.000***)	(0.000***)	(0.000***)
Event Driven Q4	2.07	6.45	0.091		−4.94%		0.719
					(0.005***)		(0.000***)
Global Macro Q1	8.49	11.09	0.216	0.59	−0.85%	−4.68%	0.230
				(0.558)	(0.734)	(0.215)	(0.001***)
Global Macro Q4	7.09	8.40	0.237		3.82%		0.473
					(0.178)		(0.000***)
Equity Market Neutral Q1	3.72	4.57	0.232	0.56	1.91%	1.07%	0.113
				(0.572)	(0.432)	(0.710)	(0.033**)
Equity Market Neutral Q4	3.21	2.78	0.329		0.85%		0.051
					(0.581)		(0.166)
Panel C							
Convertible Arbitrage Q1	6.18	1.80	0.960	5.42	3.90%	1.84%	0.770
				(0.000***)	(0.000***)	(0.735)	(0.000***)
Convertible Arbitrage Q4	5.57	13.00	0.120		2.05%		0.600
					(0.706)		(0.083*)
CTA Q1	5.66	11.36	0.140	1.21	4.66%	6.05%	0.350
				(0.224)	(0.039**)	(0.104)	(0.000***)

Table 3 *continued*

Style portfolios	*Annual excess return (%)*	*Annual volatility (%)*	*Sharpe ratio*	$\tilde{d}^{\star m}$*(sign.)*	*Alpha*	*Alpha spread (sign.)*	R^2
CTA Q4	0.97	11.39	0.020		−1.40% (0.639)		0.140 (0.000★★★)
Event Driven Q1	6.37	2.47	0.730	5.90 (0.000★★★)	3.63% (0.000★★★)	8.47% (0.000★★★)	0.450 (0.000★★★)
Event Driven Q4	1.14	6.73	0.050		−4.85% (0.010★★)		0.400 (0.000★★★)
Global Macro Q1	7.31	5.22	0.390	1.56 (0.119)	3.39% (0.061★)	1.11% (0.713)	0.570 (0.000★★★)
Global Macro Q4	4.93	6.83	0.210		2.28% (0.354)		0.080 (0.000★★★)
Equity Market Neutral Q1	4.93	1.45	0.960	4.22 (0.000★★★)	2.49% (0.000★★★)	0.96% (0.603)	0.490 (0.000★★★)
Equity Market Neutral Q4	2.73	3.31	0.240		1.53% (0.379)		0.690 (0.015★★)

This table presents average annual excess return, volatility, the Sharpe ratio and alphas for each quartile style portfolio (Q1 indicates top-quartile portfolio and Q4 bottom-quartile portfolio) from the 2000–2006 holding period. In addition, the performance differences of top- and bottom-quartile portfolios are shown by the Ledoit-Wolf test[32](that is, $\tilde{d}^{\star m}$-statistics) and by alpha spread test. The holding period alphas are estimated using 9-factor OLS regression over the stacked time series of portfolio returns. Panel A shows the results based on the use of standard 9-factor OLS alphas as the selection criterion for quartile portfolios. Correspondingly, Panel B shows the results based on 9-factor Bayesian alphas as selection period performance metrics, and Panel C the results based on the selection period Sharpe ratio. Adjusted R^2s of each regression are presented in the last column (significance levels are in parentheses. ★★★, ★★and ★ denote significance at the 1 per cent, 5 per cent and 10 per cent level, respectively).

cross-section tests of alphas when the full sample of equity market neutral funds is used as a basis of analysis. Thus, the results would indicate that for the sample employed, the persistence in equity market neutral style is concentrated on middle performers, parallel to the overall results of Capocci and Hübner[39] and Capocci *et al.*[7]

A comparison of the cross-sectional results for event-driven style with the corresponding results from top- and bottom-portfolio performance difference tests reveals a contrasting phenomenon to that observed among equity market neutral funds; while cross-sectional results do not indicate persistence when alphas are employed as input variables, the alpha spreads between top- and bottom-quartile portfolios based on both OLS and Bayesian alpha rankings are highly significant. This result indicates that for the sample employed, the persistence within event-driven style is concentrated on top- and bottom-quartiles (consistently with the overall findings of Baquero *et al*[6]).

The alpha spread of top and bottom quartiles for the stacked holding period is statistically significant for two fund styles when the Bayesian alphas are used as a selection criterion (that is, for event driven and CTA styles; see Table 3, Panel B). Correspondingly, when quartile portfolios are formed on the basis of the 9-factor OLS alpha or the Sharpe ratio, the alpha spread is significant only for event-driven style. However, *t*-statistics are also quite close to the level of statistical significance for CTA style, particularly when portfolio selection is based on the Sharpe ratio (significance level in this case is 10.4 per cent; see Panel C). For global macro and equity market neutral styles, the results of the alpha spread tests are somewhat mixed, but none are statistically significant.

When the holding period performance is evaluated on the basis of the Sharpe ratio, the results are unanimous on the significant persistence within event-driven style. The Ledoit-Wolf test statistic is also significant for CTA style when either OLS or Bayesian alphas are employed as selection criteria. For convertible arbitrage style, the Sharpe ratio differences between top- and bottom-quartile portfolios are highly significant when fund portfolio selection is based either on the Sharpe ratio or the Bayesian alpha, but insignificant based on the OLS alpha criterion. It is also noteworthy that the use of the Sharpe ratio as a selection criterion leads to positive and significant top-quartile portfolio alphas in every fund style examined (Panel C).

The overall results give some evidence that the use of the empirical Bayesian method in determining selection period alphas slightly enhances the prediction power of past alphas within the hedge fund sample employed.[40] Thus, it would be an interesting extension to study whether these findings hold also for larger samples of hedge funds. Moreover, applying the Bayesian approach has so far been applied only for the selection period (for example see Huij and Verbeek,[14] Kosowski *et al*[9]). However, as the empirical Bayesian method could be characterized as a shrinkage procedure of regression estimates (for example, see Huij and Verbeek[41] for further details), it would be justified to treat both periods equally and apply the same procedure for performance evaluation of the holding period. In this way, the potential benefits of the Bayesian approach compared to those of the standard-frequentist approach could be revealed more transparently. This kind of approach can be motivated also by the fact that the holding period performance is often determined based on shorter time-series of returns than is the selection period performance. Given that one argument for using the Bayesian approach is to alleviate estimation problems stemming from a short sample, it would be even more justified to employ the Bayesian approach in determining performance for the shorter of evaluation periods. Although this issue is beyond the scope of this article, it provides an interesting topic for further research.

The overall results are consistent with those of Harri and Brorsen[4] in the sense that performance persistence within event-driven style is found in both studies. In another sense, the results are contrasting because unlike Harri and Brorsen, we do not find evidence of persistence within global macro style. Parallel to the results of Capocci *et al*[7] or Kosowski *et al*,[9] our cross-sectional results show persistence within equity market neutral style when the same evaluation criterion is used for both selection and holding periods, though significant persistence based on performance difference tests between top- and bottom-quartile portfolios within this style is found only when the Sharpe ratio is employed as a performance metric for both periods. However, the differences in results between studies may stem from several sources: the significance and even the direction of the results often depend on sample data, test period, and the length of selection and holding periods employed, in addition to the methodology, as shown in this article. The recent results of Gregoriou *et al*[8] support this conclusion.

CONCLUSIONS

We applied two complementary methodologies in detecting the performance persistence of five different hedge fund styles. First, we determined the prediction power of the selection period performance on the holding period performance by means of cross-sectional regression. Second, we formed top-and bottom-quartile portfolios based on selection period performance and tested their holding period performance difference based on both the Ledoit-Wolf test and the alpha spread test. In addition, we compared the results obtained by using model-free performance metrics to those obtained by using alphas of multifactor models. The 3-year selection period performance was evaluated with the Sharpe ratio, 9-factor OLS alpha and the corresponding Bayesian alpha, and the 1-year holding period performance was determined based on the first two of these.

Our results showed that both the degree and existence of performance persistence vary among hedge fund styles, and, in addition, depend on both methodology and performance metric employed. In comparison to the performance metrics that are based on factor models, model-free performance metrics such as the Sharpe ratio turned out to be more sensitive in detecting performance persistence within the sample employed. The evidence of this was particularly strong based on cross-sectional regression tests. This finding is not explained by the inability of the Sharpe ratio to cope with the non-normal return distributions, as at the outset of the study we formed quartile portfolios based on several other model-free performance metrics that take account of distributional non-normalities (including the Sortino ratio, modified Sharpe ratio and adjusted Sharpe ratio), and similar to the recent results of Eling and Schuhmacher,[26] they all led to virtually identical portfolio compositions. Therefore, the main finding also holds for these alternative model-free performance metrics. However, one should be cautious in generalizing these results, as for some unknown reason the averages of both skewness and kurtosis were relatively low for this random sample compared to those reported in previous hedge fund studies. It is also noteworthy that the use of the Sharpe ratio as a selection criterion led to positive and significant top-quartile portfolio alphas in every fund style examined, indicating that the finding is somewhat robust to the choice of the holding period performance measure, and thus not a consequence of model misspecification bias. In addition, the overall results show that the Bayesian alphas are somewhat better predictors of future performance than are their OLS counterparts.

The strongest evidence of performance persistence was found among event-driven funds, for which 10 out of 12 persistence tests performed indicate significant results. Based on the cross-sectional regression tests of the Sharpe ratios, persistence is also strong among equity market neutral funds. The results between these two styles deviated from each other in that persistence within event-driven style is concentrated on top and bottom quartiles whereas among equity market neutral funds it is concentrated on middle performers. Although our sample includes only one directional fund style, our results are in line with those of Gregoriou *et al*,[8] who report somewhat stronger persistence within non-directional styles (such as convertible arbitrage, equity market neutral or event driven) than within directional styles (such as global macro).

ACKNOWLEDGEMENTS

We thank Tomi Seppälä for assisting in programming the empirical Bayesian procedure and Joop Huij for his helpful comments. We are indebted to TASS and RAM Partners Ltd. (particularly Ville Toivakainen) for providing us with the data.

REFERENCES AND NOTES

1 Pätäri, E.J. (2008a) A closer look at performance persistence of mutual funds. *Journal of Performance Measurement* 12(3): 38–46.
2 Brown, S.J., Goetzmann, W.N. and Ibbotson, R.G. (1999) Offshore hedge funds: Survival and performance 1989–95. *Journal of Business* 72(1): 91–117.
3 Agarwal, V. and Naik, N.Y. (2000a) Multi-period performance persistence analysis of hedge funds. *Journal of Financial and Quantitative Analysis* 35(3): 327–342.
4 Harri, A. and Brorsen, B. (2004) Performance persistence and the source of returns for hedge funds. *Applied Financial Economics* 14(2): 131–141.
5 Agarwal, V. and Naik, N.Y. (2000b) On taking the 'alternative' route: Risks, rewards, style and performance persistence of hedge funds. *Journal of Alternative Investments* 2(4): 6–23.
6 Baquero, G., ter Horst, J. and Verbeek, M. (2005) Survival, look-ahead bias, and persistence in hedge fund performance. *Journal of Financial and Quantitative Analysis* 40(3): 493–517.
7 Capocci, D., Corhay, A. and Hübner, G. (2005) Hedge fund performance and persistence in Bull and Bear markets. *European Journal of Finance* 11(5): 361–392.
8 Gregoriou, G.N., Hübner, G., Papageorgiou, N. and Rouah, F.D. (2007) Fund of funds versus simple portfolios of hedge funds: A comparative study of persistence in performance. *Journal of Derivatives and Hedge Funds* 13(2): 88–106.
9 Kosowski, R., Naik, N.Y. and Teo, M. (2007) Do hedge funds deliver alpha? A Bayesian and Bootstrap analysis. *Journal of Financial Economics* 84(1): 229–264.
10 Baks, K.P., Metrick, A. and Wachter, J. (2001) Should investors avoid all actively managed mutual funds? A study in Bayesian performance evaluation. *Journal of Finance* 56(1): 45–85.
11 Pástor, L. and Stambaugh, R.F. (2002) Mutual fund performance and seemingly unrelated assets. *Journal of Financial Economics* 63(3): 315–349.
12 Bollen, N.P. and Busse, J.A. (2005) Short-term persistence in mutual fund performance. *Review of Financial Studies* 18(2): 569–597.
13 Busse, J.A. and Irvine, P.J. (2006) Bayesian alphas and mutual fund persistence. *Journal of Finance* 61(5): 2251–2288.

14 Huij, J. and Verbeek, M. (2007) Cross-sectional learning and short-run persistence in mutual fund performance. *Journal of Banking and Finance* 31(3): 973–997.
15 Jones, C.S. and Shanken, J. (2005) Mutual fund performance with learning across funds. *Journal of Financial Economics* 78(3): 507–522.
16 Fung, W. and Hsieh, D. (2004) Hedge fund benchmarks: A risk based approach. *Financial Analyst Journal* 60(5): 65–80.
17 The same additional factor is employed by Harri and Brorsen,[4] Capocci *et al*[8] and Gregoriou *et al*[7].
18 For the sample data, the adjusted R^2s of the 9-factor models are on average twice those of the Fung-Hsieh 7-factor models. For the 3-year selection period, the average of the adjusted R^2s are 0.316 for the 9-factor model and 0.186 for the 7-factor models, and for the 1-year holding period they are 0.346 and 0.150, respectively. The results are also parallel within each of the five hedge fund styles examined.
19 Fung, W. and Hsieh, D. (2001) The risk in hedge fund strategies: Theory and evidence from trend followers. *Review of Financial Studies* 14(2): 313–341.
20 Lo, A.W. (2002) The statistics of Sharpe ratios. *Financial Analysts Journal* 58(4): 36–52.
21 Brooks, C. and Kat, H.M. (2002) The statistical properties of hedge fund index returns and their implications for investors. *Journal of Alternative Investments* 5(2): 26–44.
22 Gregoriou, G.N. and Gueyie, J.P. (2003) Risk-adjusted performance of funds of hedge funds using a modified Sharpe ratio. *Journal of Wealth Management* 6(3): 77–83.
23 Mahdavi, M. (2004) Risk-adjusted return when returns are not normally distributed: Adjusted Sharpe ratio. *Journal of Alternative Investments* 6(4): 47–57.
24 Sharma, M. (2004) A.I.R.A.P. – Alternative RAPMs for alternative investments. *Journal of Investment Management* 2(4): 106–129.
25 Morton, D.P., Popova, E. and Popova, I. (2006) Efficient fund of hedge funds construction under downside risk measures. *Journal of Banking and Finance* 30(2): 503–518.
26 Eling, M. and Schuhmacher, F. (2007) Does the choice of performance measure influence the evaluation of hedge funds? *Journal of Banking and Finance* 31(9): 2632–2647.
27 Ferruz, L., Sarto, J.L. and Andreu, L. (2008) Do asymmetric risk metrics influence performance persistence? *Journal of Derivatives and Hedge Funds* 14(1): 42–49.
28 Ferruz, L., Pedersen, C. and Sarto, J.L. (2006) Performance metrics for Spanish investment funds. *Derivatives Use, Trading & Regulation* 12(3): 219–227.
29 Pätäri, E.J. (2008b) Comparative analysis of total risk-based performance measures. *Journal of Risk* 10(4): 69–112, The author reports remarkable differences in risk rankings of equity mutual funds, when more extreme partial scale measures of downside risk are employed, but finds corresponding performance rank orders far more stable. When more conservative downside risk metrics, such as semi-standard deviation, absolute semi-deviation or Gini's mean difference, are employed, the resulting performance rankings are nearly identical to those based on the Sharpe ratio.
30 Favre, L. and Signer, A. (2002) The difficulties of measuring the benefits of hedge funds. *Journal of Alternative Investments* 5(1): 21–25.
31 Hu, W. and Maddala, G. (1994) Estimation and Prediction Problems in Dynamic Heterogeneous Panel Data Models. Ohio State University, Working Paper.
32 Because of the complexity of the test procedure and space limitations, we do not describe the Ledoit-Wolf test in detail, but recommend that the interested reader see the original article (corresponding programming code is freely avalaible at http://www.iew.uzh.ch/chairs/wolf/team/wolf/publications.html#7): Ledoit, O. and Wolf, M. (2008) Robust performance hypothesis testing with the Sharpe ratio. forthcoming in *Journal of Empirical Finance* 15(5): 850–859.
33 Brown, S.J., Goetzmann, W.N. and Park, J. (2001) Careers and survival: Competition and risk in the hedge fund and CTA industry. *Journal of Finance* 56(5): 1869–1886.
34 Gregoriou, G.N. (2002) Hedge fund survival lifetimes. *Journal of Asset Management* 2(3): 237–252.
35 Amin, G.S. and Kat, H.M (2003) Welcome to the dark side: Hedge fund attrition and survivorship bias over the period 1994–2001. *Journal of Alternative Investments* 6(1): 57–73.
36 Requested citation: http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-Fac.xls(The same PTFS index data are also employed by Fung and Hsieh[19] and Kosowski *et al*,[9] for example).
37 For example, see Carhart, M.M. (1997) On persistence in mutual fund performance. *Journal of Finance* 52(1): 57–82, As noted by the author, using the same performance measure for both selection and holding periods raises model misspecification bias as follows: if the factor-mimicking portfolios impose risk premia too high or too low, funds with consistent factor model loadings will show persistent factor model performance. A similar problem is incurred if there is an omitted factor in the model.
38 As an additional test, we perform the alpha spread test for those top and bottom portfolios that consist only of five best and worst funds within each style. The results are parallel to those based on quartile portfolios: the strongest persistence within event driven style and the

weakest within global macro style. The confidence levels of the significant results are higher when portfolios are formed on the Bayesian alphas instead of on OLS alphas. Moreover, alpha spreads are generally more significant between top and bottom quartile portfolios than they are between top-five and bottom-five portfolios (the detailed results of these tests are available from the authors upon request).

39 Capocci, D. and Hübner, G. (2004) Analysis of hedge fund performance. *Journal of Empirical Finance* 11(1): 55–89.

40 Altogether, we performed 22 different persistence tests using each type of alpha as prediction criteria. Evidence of persistence was found in 10 out of 22 cases when the selection period performance metric was Bayesian alpha, and in only five out of 22 cases when using OLS alphas as the corresponding measure (see Tables 2 and 3).

41 Huij and Verbeek, M. (2003) Evaluating Mutual Fund Performance and Its Persistence Using Shrinkage Estimators. Erasmus University Rotterdam, The Netherlands, Working Paper.