Open

Original Article

A simulation-based methodology for evaluating hedge fund investments

Received (in revised form): 16th November 2015

Marat Molyboga

is the Chief Risk Officer and Director of Research at Efficient Capital Management, LLC and a member of the Executive Committee. He is also an Adjunct Professor of Finance at the Illinois Institute of Technology Stuart School of Business. He began his career at Efficient in 2001 as a Research Analyst. Beginning in 2002, he consulted for Petra Intraday Trading Systems as its Senior Researcher and served as its president from 2004 until 2006 when he joined Research Department of Efficient as a Senior Research Analyst.

Christophe L' Ahelec

is a Portfolio Manager in the Alternative Investments group at Ontario Teachers' Pension Plan Board where he is responsible for the portfolio construction and risk management of the hedge fund portfolio. Before that he was a quantitative analyst and assistant portfolio manager at Mignon Genève SA/Alpstar Asset Management, in Switzerland, where he took part in the creation and portfolio management of a European systematic market neutral fund. He is a graduate engineer in Finance and Applied Mathematics from the Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble, France and holds the Chartered Financial Analyst® designation.

Correspondence: Marat Molyboga, Efficient Capital Management, Director of Research, 4355 Weaver Parkway, Warrenville, IL 60555, USA.
E-mail: molyboga@efficient.com

ABSTRACT This article introduces a large scale simulation framework for evaluating hedge funds' investments subject to the realistic constraints of institutional investors. The method is customizable to the preferences and constraints of individual investors, including investment objectives, performance benchmarks, rebalancing period and the desired number of funds in a portfolio and can incorporate a large number of portfolio construction and fund selection approaches. As a way to illustrate the methodology, we impose the framework on a subset of hedge funds in the managed futures space that contains 604 live and 1323 defunct funds over the period 1993–2014. We then measure the out-of-sample performance of three hypothetical risk-parity (RP) portfolios and two hypothetical minimum risk portfolios and their marginal contributions to a typical 60–40 portfolio of stocks and bonds. We find that an investment in managed futures improves an investor's performance regardless of portfolio construction methodology and that equal risk approaches are superior to minimum risk portfolios across all performance metrics considered in the study. Our article is relevant for institutional investors in that it provides a robust and flexible framework for evaluating hedge fund investments given the specific preferences and constraints of individual investors.
Journal of Asset Management (2016) **17,** 434–452. doi:10.1057/jam.2016.3;
published online 24 March 2016

Keywords: large-scale simulation framework; hedge funds; optimal portfolios; risk parity; risk-based allocations

The online version of this article is available Open Access

INTRODUCTION

The hedge fund industry represented about US $3 trillion in assets under management (AUM) during the first quarter of 2015 according to the BarclayHedge Group. Therefore, hedge funds represent a significant portion of the portfolios of institutional investors with direct investments of $2.5 trillion and an additional $500 billion allocated through funds of funds. While there is a rich literature on quantitative approaches to portfolio construction, most studies fail to appropriately account for investment practices and, therefore, are not directly applicable for institutional investors who have their own unique set of investment constraints and preferences. Molyboga *et al* (2015), henceforth MBB, suggest that institutional investors cannot directly benefit from academic studies on hedge fund performance persistence because the studies: (i) ignore the delay in hedge fund reporting, thus relying on information that is not available at the time of investment decisions,[1] (ii) consider funds that have AUM that are too small for institutional investors, (iii) include funds with very short track records, and (iv) assess portfolios with too many constituent funds to be practical.[2] MBB apply a one-month lag parameter to account for the reporting delay, impose AUM and track record length requirements, and introduce a simulation framework that limits the number of funds within a portfolio and then evaluate out-of-sample performance using a stochastic dominance framework.

This article uses a simulation framework similar to MBB but applies it to the evaluation of portfolio construction approaches subject to real life constraints and uses additional performance metrics that evaluate the marginal portfolio contribution of a hedge fund portfolio to an investor's original portfolio. This study intentionally uses a few commonly used measures of performance to illustrate that the framework is not limited to a single measure. The framework is customizable to the preferences and constraints of individual investors regarding rebalancing periods and the desired number of funds in a portfolio and can incorporate a large number of portfolio construction and fund selection approaches. The methodology produces implementable results because it explicitly accounts for the hedge fund reporting delay reported in MBB and applies an in-sample/out-of-sample framework that incorporates common investment constraints when creating and rebalancing portfolios. The framework imposes the standard requirements of institutional investors regarding track record length and the amount of AUM. It also limits the number and turnover of funds in the portfolio by assuming that the institutional investor selects a discrete number of funds that stay in the portfolio until they no longer satisfy selection criteria.[3] The methodology utilizes a simulation framework to account for a large number of feasible portfolio constituents in each period.

We evaluate out-of-sample performance with several commonly used measures of standalone performance and marginal portfolio contribution.[4] Standalone performance measures include annualized return, Sharpe and Calmar ratios, maximum drawdown[5] and the *t*-statistic of α with respect to the Fung-Hsieh (2001) five-factor model. We measure marginal portfolio contribution by evaluating the improvement in Sharpe and Calmar ratios[6] by replacing a modest 10 per cent of the original investor's portfolio with a 10 per cent allocation to a simulated hedge fund portfolio. In this article, we consider a standard 60–40 portfolio of stocks and bonds as the original portfolio, but the framework is flexible to the choice of investor benchmark.

Standard statistical techniques are inappropriate for the evaluation of out-of-sample performance since simulation results are not independent, driven rather by the overlap in portfolio constituents across simulations. We apply the bootstrapping methodology of Efron (1979) and Efron and Gong (1983) to estimate the sampling properties of the test results and draw statistical inferences about the relative performance of portfolio methodologies. Opdyke (2007) introduces an analytic formula for the asymptotic distribution of Sharpe ratios under very general conditions that include non-normal distributions, time-varying volatility and serial correlations. This approach is particularly powerful when applied to a single return series such as in asset allocation studies that use single indices for each asset class. For example, studies by Kat (2004), Lintner (1996), Abrams *et al* (2009) and Chen *et al* (2005) that demonstrate positive contribution of managed futures to traditional portfolios can directly benefit from utilizing the methodology introduced in Opdyke (2007). By contrast, the simulation methodology of this article produces many time-series and we select a boostrapping approach because it can be applied to any performance measure while accounting for lack of independence in simulation results.

We impose the framework with 10 000 simulations on a data set of 604 live and 1323 defunct Commodity Trading Advisors (CTAs) over the period 1993–2014. CTAs, a subset of hedge funds that has grown exponentially over the past 35 years,[7] is known for its historically strong performance during times of market crisis, notably the Financial Crisis of 2008, and, therefore, serves as a particularly interesting subset of hedge funds from a portfolio diversification perspective. We evaluate several popular risk-based approaches that include two minimum risk and three RP methods. While the approaches we consider are commonly used by both practitioners and academics, they are only a few of the portfolio construction approaches that can be evaluated within the framework. The methodology can be extended to a large number of quantitative portfolio construction approaches.

Our results are striking because an investment in CTAs improves performance regardless of the choice of the portfolio construction approach. For the out-of-sample period between January 1999 and December 2014, a 10 per cent allocation to managed futures improves the Sharpe ratio of the original 60–40 portfolio of stocks and bonds from 0.376 to 0.399–0.416 on average, depending on the portfolio construction methodology employed. Similarly, the Calmar ratio improves from 0.092 to 0.100–0.108 on average. Blended portfolios have higher Sharpe ratios in at least 89 per cent of simulations and higher Calmar ratios in at least 89.5 per cent of simulations. Our findings are consistent with Kat (2004), Lintner (1996), Abrams *et al* (2009) and Chen *et al* (2005) that report a positive contribution of managed futures to traditional portfolios.

Minimum risk portfolios perform the worst for all performance metrics. For example, their average Sharpe ratios are between 0.299 and 0.304, significantly lower from both an economic and statistical perspective than the 0.319 average Sharpe ratio of the random portfolios. By contrast, equal risk methodologies deliver superior average Sharpe ratios of 0.342–0.362. Our results are consistent with DeMiquel *et al* (2009) who find that an equal notional allocation (EN), which we consider an equal-risk approach, is superior to a minimum variance allocation (MV), which we consider a minimum risk approach.

We have performed a sub-sample analysis to evaluate the marginal contribution of a 10 per cent allocation to managed futures during periods of relatively poor and relatively good performance. A 60–40 portfolio produced a Sharpe ratio of −0.09 and a Calmar ratio of −0.025 during the relatively poor period between 1999 and 2008. During the exceptional period between 2009 and 2014, the benchmark portfolio delivered a Sharpe ratio of 1.12 and a Calmar ratio of 0.77.[8] During the 10-year period between 1999 and 2008, all portfolio construction approaches would have added value as measured in terms of average Sharpe and Calmar ratios with the equal-risk portfolios producing the best results.

 © 2016 Macmillan Publishers Ltd. 1470-8272

During the 6-year period between 2009 and 2014, the blended portfolios have approximately the same average Sharpe ratios and slightly better Calmar ratios than the benchmark portfolio with the minimum risk approaches delivering the best results. The sensitivity of relative results to an evaluation window is common. For example, Anderson *et al* (2012) compare four investment strategies and find that the specific start and end dates of a backtest can have a material impact on the results. This study introduces a methodology that can be used for the evaluation of portfolio construction approaches with real-life constraints and is strengthened by the sub-sample analysis.

Our findings and methodology are relevant for institutional investors who might consider investing, or who are already currently invested, in hedge funds and managed futures because the framework can be customized to the specific preferences and constraints of investors to maximize the benefits of hedge fund portfolios.

The remainder of the article is organized as follows: The next section describes the data and accounts for biases; the subsequent section discusses the risk-based approaches and introduces the large-scale simulation framework; the section after that presents empirical out-of-sample results; and the final section concludes.

DATA

There are several commonly used CTA databases: BarclayHedge; CISDM (formerly the MAR database); Lipper (formerly TASS); and Eurekahedge. Joenvaara *et al* (2012) perform a comprehensive study of publicly available databases of hedge fund returns and report that Barclay Hedge provides the highest quality data out of the databases considered. Moreover, the BarclayHedge database is the largest publicly available database of CTAs with 1013 active and 3660 defunct funds over the period from December 1993 to December 2014. Therefore, we use BarclayHedge for this study as it is the most comprehensive and highest quality publicly available database of CTA returns.

We perform a number of filtering steps to ensure data quality and limit the scope of the study to the funds that would be appropriate for institutional investors who are interested in making direct investments. We explicitly account for the survivorship, backfill, incubation and liquidation biases that are common within CTA and hedge fund databases.[9] We include the graveyard database that contains defunct funds to account for the survivorship bias. The backfill and incubation biases arise because of the voluntary nature of self-reporting.[10] We use a combination of two approaches to mitigate these biases. The first methodology, suggested by Fama and French (2010), limits the tests to those funds that managed at least $10 million in AUM normalized to December 2014 values. Once a fund reaches the AUM minimum, it is included in all subsequent tests to avoid creating selection bias. Unfortunately, many CTAs, including very successful and established ones, originally reported only net returns for an extended period of time before their initial inclusion of AUM data. Using Fama and French (2010) methodology exclusively would completely eliminate large portions of valuable data for such funds. To include this data, we apply the technique suggested by Kosowski *et al* (2007), which eliminates only the first 24 months of data for such funds. We use the liquidation bias estimate of 1 per cent as suggested in Ackermann *et al* (1999). After accounting for the biases, our data set includes 604 live and 1323 defunct funds for the period between December 1995 and December 2014.

We use the Fung-Hsieh five factor model of primitive trend following systems, introduced in Fung and Hsieh (2001), as benchmarks in measuring the performance of CTA portfolios. The factors include PTFSBD (bonds), PTFSFX (foreign exchange), PTFSCOM (commodities), PTFSIR (interest rates) and PTFSSTK (stocks) while the 3-month Treasury bill (secondary market rate) series with ID TB3MS from the Board of Governors of the Federal Reserve System serves as a proxy for the risk-free rate. Table 1 reports summary statistics and tests of normality, hetero-scedasticity and serial correlations in CTA returns by strategy and current status.

Anson (2011) suggests that the 60–40 portfolio of stocks and bonds represents a typical starting point for a US institutional investor. In this article, this blend is constructed using the S&P 500 Total Return index and the JPM Global Government Bond Index. Table 2 reports the annualized excess return, standard deviation, maximum drawdown, Sharpe ratio and Calmar ratio of the 60–40 portfolio for 1999–2014. Over this time period, the portfolio delivered a Sharpe ratio of 0.376 and a Calmar ratio of 0.092.

Figure 1 shows the performance of the portfolio from January 1999 to December 2014.

Although the 60–40 portfolio of stocks and bonds has been used extensively in the literature as a benchmark portfolio, the framework is flexible and can incorporate any investor-specific portfolio as a benchmark.

METHODOLOGY

In this section, we define the risk-based approaches considered in this study. Then we introduce a large-scale simulation framework with real-life constraints used to generate out-of-sample portfolio returns. Finally, we describe the performance metrics used to compare out-of-sample results.

Review of risk-based approaches

In this article, we evaluate two minimum risk and three equal-risk (or RP) approaches.[11] While the approaches we consider are commonly used by practitioners and academics, they are used merely as examples of portfolio construction approaches that can be evaluated within the framework. The methodology can be extended to a large number of quantitative portfolio construction approaches. Minimum risk portfolios include the MV approach with non-negative constraints documented in Jagannathan and Ma (2003) and a minimum semi-standard deviation (MDEV) approach that is similar to the MV approach but only considers negative returns. Equal-risk or RP, approaches include an EN approach, which is a naïve diversification 1/N method praised in DeMiquel *et al* (2009) and criticized in Kritzman *et al* (2010), an EVA approach highlighted in c (2012) and the classical risk parity (RP) approach extensively discussed in Maillard *et al* (2010), Clarke *et al* (2013) and Qian (2013). We apply a random portfolio selection approach (Random) that serves as a benchmark in evaluating the RP approaches. The approaches are evaluated using a large-scale simulation framework with real life constraints.

Large scale simulation framework

In this, we utilize a modification of the large-scale simulation framework with real life constraints introduced in MBB. MBB apply the framework to evaluate persistence in hedge fund managers' performance and compare equally weighted portfolios of funds that rank in the top quintile based on the *t*-statistic of α with respect to a CTA benchmark (restrictive fund selection) against those of all available funds (random fund selection). By contrast, this article does not impose any ranking but rather focuses on the impact of choice of portfolio construction methodology on performance. The out-of-sample period is between January 1999 and December 2014, the longest out-of-sample backtesting period in CTA empirical research. The framework uses 10 000 simulations and a lag of one month to account for the delay in the performance reporting of CTAs.[12] Below we describe a single run of the

 © 2016 Macmillan Publishers Ltd. 1470-8272

Table 1: Summary statistics and tests of normality, heteroscekasticity and serial correlation in CTA returns

	Number of funds	*Mean*				*Test of normality*	*Test of heteroskedasticity*	*Test of autocorrelation*
		α (%)	*t-statistics of α*	*Kurtosis*	*Skewness*	*Funds with Jarque-Bera P<0.1 (%)*	*Funds with Breusch Pagan P<0.1 (%)*	*Funds with Ljung-Box P<0.1 (%)*
All Funds	1927	0.43	0.91	4.27	0.09	45	24	21
By Strategy:								
Arbitrage	24	0.05	0.33	6.33	−0.13	75	17	13
Discretionary	34	−0.06	0.02	4.30	−0.10	41	12	15
Fundamental – Agricultural	44	0.46	0.62	5.71	0.41	57	16	27
Fundamental – Currency	97	0.50	0.97	4.28	0.23	52	19	20
Fundamental – Diversified	105	0.38	0.95	4.13	0.15	51	21	11
Fundamental – Energy	23	0.03	0.25	4.60	0.26	52	4	9
Fundamental – Financial/Metals	79	0.24	0.77	4.70	0.12	52	9	14
Fundamental – Interest Rates	12	−0.10	−0.26	3.21	−0.06	25	42	33
Option Strategies	88	−0.16	0.04	7.21	−0.43	80	43	22
Stock Index	85	0.14	0.57	4.35	0.06	41	25	15
Stock Index,Option Strategies	3	−0.86	−1.88	6.43	−1.07	67	33	33
Systematic	39	0.41	−0.53	4.07	0.09	46	13	23
Technical – Agricultural	9	−0.50	−0.72	4.29	0.26	67	11	11
Technical – Currency	203	0.35	0.74	4.18	0.25	46	14	22
Technical – Diversified	714	0.67	1.28	3.84	0.09	38	28	22
Technical – Energy	4	−0.54	−0.46	3.89	−0.06	50	0	25
Technical – Financial/Metals	214	0.37	0.89	3.86	0.03	34	23	20
Technical – Interest Rates	11	0.46	1.37	3.37	−0.12	36	55	9
Other	139	0.40	1.10	4.46	0.07	50	25	27
By current status:								
Live funds	604	0.71	1.51	4.26	0.11	43	34	20
Dead funds	1323	0.30	0.62	4.27	0.08	45	19	21

This table reports statistical properties of fund returns and residuals by strategy and current status. Column one presents the number of funds in each category. Columns two and three report the cross-sectional mean of the Fung-Hsieh (2001) five-factor model monthly α and the t-statistic of α. Columns four and five display the cross-sectional mean of kurtosis and skewness of fund residuals. Column six reports the percentage of funds for which the null hypothesis of normal distribution is rejected by the Jarque-Bera test. Column seven reports the percentage of funds for which the null hypothesis of homoskedasticity is rejected by the Breusch Pagan test. Column eight reports the percentage of funds for which the null hypothesis of zero first-order autocorrelation is rejected by Ljung-Box test. All tests are applied to fund residuals and the P-value is set at 10% level.

simulation framework and then show how simulation results are evaluated.

A single run of the simulation framework

The in-sample/out-of-sample framework mimics the actions of an institutional investor who makes allocation decisions at the end of each month. The first decision is made in December 1998. Owing to the delay in CTA reporting, the investor has return information only through November 1998; thus, the investor considers all funds that have a complete set of monthly returns between December 1995 and November 1998. First, the investor eliminates all funds in the bottom quintile of AUM among the funds considered. This relative AUM threshold is more appropriate than the fixed AUM approach commonly used in the literature (for example, Kosowski *et al* (2007) use a fixed AUM level of $20 million) because the average level of AUM has increased substantially over the last 20 years. Then the investor randomly chooses 5 funds from the remaining pool of CTAs and allocates to them using the five risk-based approaches and a random portfolio allocation. Monthly returns are recorded for each portfolio construction approach for January 1999 using the liquidation bias adjustment for funds that liquidate during the month. At the end of January 1999, the pool of CTAs is updated and defunct constituents of the original portfolio are randomly replaced with funds from the new pool. Each portfolio is then rebalanced again using the original portfolio construction methodologies.[13] The process is repeated until the end of the out-of-sample period of December 2014. A single simulation results in six out-of-sample return streams between January 1999 and December 2014 – one for each of the portfolio construction approaches.

Table 2: Performance of a 60–40 portfolio of stocks and bonds for 1999–2014

Annualized Excess Return	3.61%
Annualized Standard deviation	9.59%
Maximum Drawdown	39.29%
Sharpe ratio	0.376
Calmar ratio	0.092

This table reports annualized excess return, standard deviation, maximum drawdown, Sharpe ratio and Calmar ratio of the 60–40 portfolio of stocks and bonds for 1999–2014. The portfolio is constructed using S&P 500 Total Return index and the JP Morgan Global Government Bond Index. 3-month Treasury bill (secondary market rate) is used as a proxy for the risk free rate. Calmar is the ratio of the annualized excess return and maximum drawdown.

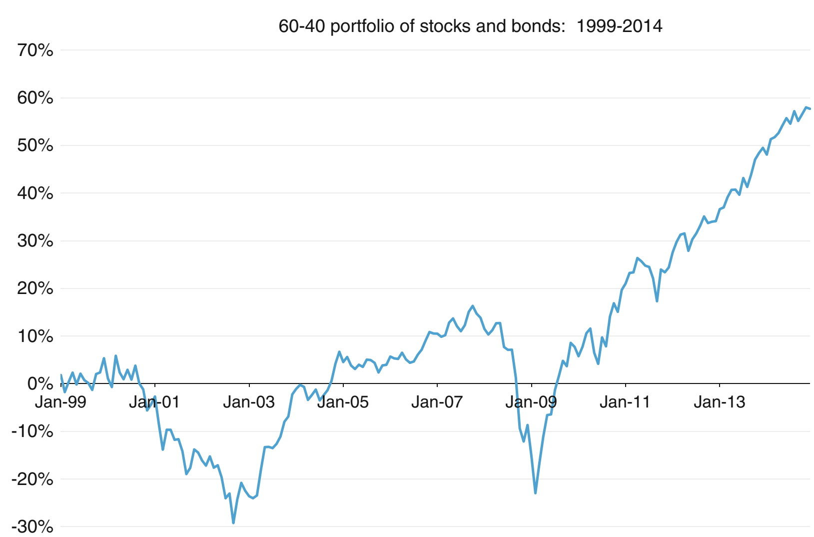

Figure 1: This figure displays performance of the 60–40 portfolio of stocks and bonds for 1999–2014. The portfolio is constructed using S&P 500 Total Return index and JP Morgan Global Government Bond index.

Performance evaluation of out-of-sample results

Out-of-sample performance is evaluated using both standalone performance metrics and measures that consider portfolio contribution benefits. Standalone performance metrics include annualized return, maximum drawdown, Sharpe ratio, Calmar ratio,[14] Fung-Hsieh α and t-statistic of α. Performance contribution is measured as the resultant difference in Sharpe ratio and Calmar ratio from replacing 10 per cent of the original portfolio of stocks and bonds with portfolios of CTA funds constructed within the simulation framework. Since each performance measure is represented by a distribution that contains 10 000 values, distributions are compared using means and medians for all measures and the percentage of positive values for Fung-Hsieh α and the percentage of positive marginal Sharpe and Calmar ratios in the performance contribution measures. Since simulations are not independent, we apply a bootstrapping procedure to draw statistical inference.

Boostrapping procedure

The bootstrapping procedure follows each steps of the simulation framework but limits the set of portfolio construction approaches to the Random portfolio methodology to which we choose to compare all other approaches.[15] Each simulation set consists of 10 000 simulations. The bootstrapping procedure includes 400 sets of simulations, a sufficient number to estimate P-values with high precision. A comparison of the performance metrics of the original simulation to the bootstrapped sets of simulations gives the P-values reported in the empirical results section.

EMPIRICAL OUT-OF-SAMPLE RESULTS

In this section, we present information about the data set used in the simulation and out-of-sample results for the period between January 1999 and December 2014 generated by the large-scale simulation framework.

Table 3: Annual statistics of commodity trading advisors

Year	*AUM threshold*	*Number of funds*
1999	8 850 000	176
2000	5 600 000	180
2001	5 020 000	186
2002	5 037 700	194
2003	9 930 000	195
2004	10 874 500	220
2005	10 423 900	237
2006	13 348 000	248
2007	12 499 700	286
2008	11 734 200	314
2009	13 422 100	337
2010	13 970 300	354
2011	13 380 000	365
2012	10 290 700	354
2013	12 295 000	336
2014	11 527 300	315

This table presents threshold level of AUM assigned at the bottom 20% level and the number of funds with at least 36 months of returns used in the study.

Table 3 reports the average AUM threshold level for each year and the average number of funds meeting that threshold. The AUM threshold represents the 20th percentile of AUM among all active fund managers with a track record of at least 36 months.

There is a significant variation in the values of the AUM threshold over time which primarily reflects changes in AUM driven by industry growth and recent performance. The 2010 threshold value of $13.97 million is almost three times as high as the $5 million threshold value in 2001. The number of funds has nearly doubled over this time period representing substantial growth in the industry.

Analysis of out-of-sample performance of CTA portfolios as standalone investments

We analyze distributions of out-of-sample returns over the complete data period using means and medians of several performance metrics. Since simulations are not independent, we use a bootstrapping

Table 4: Mean and median statistics of out-of-sample performance 1999–2014

Portfolio construction approach	*Return (%)*	*Volatility (%)*	*Sharpe*	*Calmar*	*Maximum drawdown (%)*
Panel A. Mean values					
RANDOM	3.72	11.75	0.319	0.154	28.12
EN	3.73	11.03	0.342	0.168	24.98
EVA	2.95	8.21	0.358	0.174	19.12
RP	3.13	8.66	0.362	0.176	20.40
MV	2.13	6.79	0.304	0.136	19.90
MDEV	2.10	6.80	0.299	0.134	19.91
Panel B. Median values					
RANDOM	3.67	11.62	0.317	0.135	26.81
EN	3.67	10.91	0.337	0.154	24.08
EVA	2.84	8.02	0.354	0.156	18.21
RP	3.04	8.48	0.358	0.157	19.42
MV	1.93	6.66	0.298	0.106	18.11
MDEV	1.91	6.67	0.297	0.106	18.09

This table presents mean and median values of out-of-sample performance measures for each portfolio construction approach. Performance measures include annualized excess return, annualized excess standard deviation, Sharpe and Calmar ratio (defined as annualized excess return over maximum drawdown), and maximum drawdown. Panel A reports mean values, Panel B displays median values.

methodology to draw statistical inferences about the relative performance of portfolio construction approaches.

Distributions of out-of-sample performance

Table 4 reports means and medians for the distributions of returns, volatilities, Sharpe and Calmar ratios and maximum drawdowns for each portfolio construction approach. The *P*-values are estimated using the bootstrap methodology. The superscript star indicates that the performance measure of a given portfolio approach exceeds that of the RANDOM portfolio at 99 per cent confidence level. The subscript star shows that the performance measure of a given portfolio approach is lower than that of the RANDOM portfolio at 99 per cent confidence level.

The minimum risk approaches tend to have the lowest volatilities of the portfolio methodologies considered in the study. MV and MDEV have mean volatilities of around 6.8 per cent, whereas, EVA and RP have volatilities of around 8.21 and 8.66 per cent, respectively, followed by EN and RANDOM with volatilities that exceed 11 per cent. However, the lower levels of volatility are not necessarily associated with lower drawdowns. For example, EVA has a maximum drawdown of 19.12 per cent, slightly lower than the 19.9 per cent maximum drawdown values of the minimum risk portfolios. Moreover, the minimum volatility approaches deliver low returns and risk-adjusted returns that are inferior to those of the other approaches. This finding is consistent with DeMiquel et al (2009) which documents the superior out-of-sample performance of the naïve 1/N (EN) approach relative to that of several extensions of mean-variance optimization including the MV approach. Jensen's inequality suggests the EN approach should dominate the RANDOM methodology in terms of Sharpe ratio because of the concavity of the Sharpe ratio.[16] The three equal-risk approaches have risk-adjusted performance which is superior to that of the RANDOM approach. In contrast, minimum risk approaches yield inferior results on average. Median values reported in Panel B show similar results.

While Table 4 presents mean and median values of several performance metrics, a complete evaluation of the portfolio construction methodologies should also

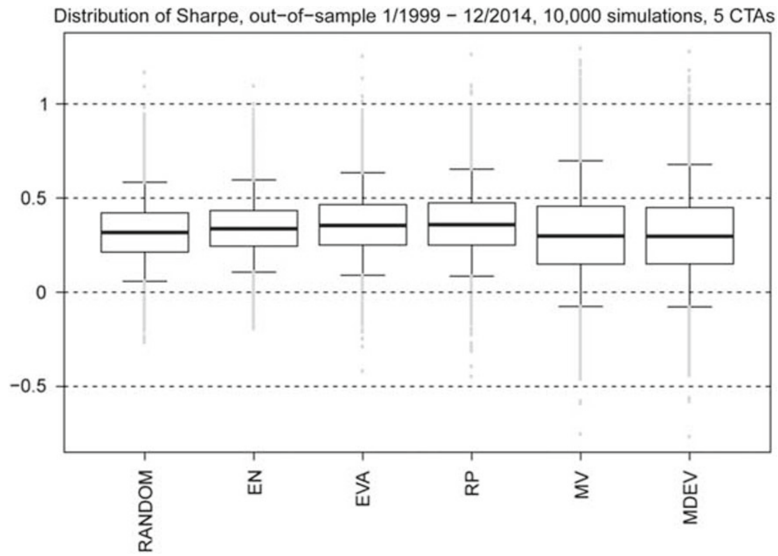

Figure 2: This figure shows distributions of the Sharpe ratios, generated using the large-scale simulation framework for the out-of-sample period between January 1999 and December 2014.

consider distributions of out-of-sample performance. Figure 2 shows the distributions of Sharpe generated by the large-scale simulation framework for each portfolio methodology.

Each distribution is visualized using a standard box and whisker plot with the box containing the middle two quartiles, the thick line inside the box representing the median of the distribution and the whiskers displayed at the top and bottom 5 per cent of the distribution. The breadth of each distribution demonstrates the key benefit of using a large-scale simulation framework. Failing to account for the role of chance and evaluating portfolio construction techniques using a single stream, which represents a single draw of the distribution, can mislead investors about the relative performance of portfolio management techniques. Since the distributions are so wide, it might seem impossible to compare them with each other. Fortunately, it is not a new problem in Quantitative Finance and Decision Theory where expected utility and stochastic dominance methodologies are applied to compare distributions. The framework is flexible and can employ utility functions and stochastic dominance to evaluate results; however, this article only considers means and medians for the sake of brevity.[17] The minimum risk approaches, MV and MDEV, have the lowest median Sharpe and exhibit relatively large left tails. The equal risk approaches seem to perform better on average than the random portfolio methodology, but it is difficult to determine whether that relative performance is statistically significant, particularly since the standard statistical techniques are inappropriate due to dependence across simulation results. Therefore, we apply a bootstrapping procedure to estimate sampling distributions of the performance measures.

The *P*-values suggest that equal risk approaches (EN, EVA, RP) dominate RANDOM portfolios based on average Sharpe and Calmar ratios at a confidence level greater than 99 per cent (in fact, none of the 400 bootstrap simulations of RANDOM portfolios deliver superior average Sharpe and Calmar ratios). By contrast, minimum risk approaches (MV, MDEV) are inferior to RANDOM portfolios in terms of average Sharpe and Calmar ratios (all 400 bootstrap simulations of RANDOM portfolios yield superior average Sharpe and Calmar ratios).[18]

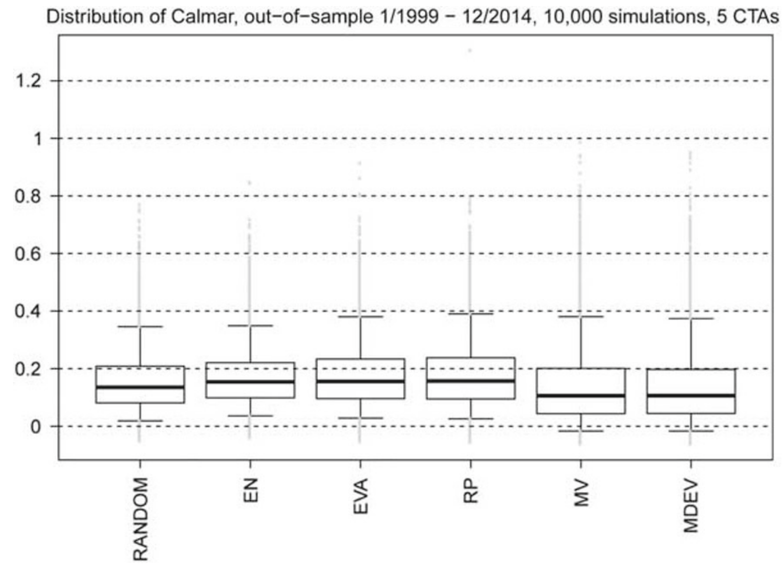

Figure 3: This figure shows distributions of the Calmar ratios, generated using the large-scale simulation framework for the out-of-sample period between January 1999 and December 2014.

Figure 3 displays the distribution of Calmar ratios.

The minimum risk approaches, MV and MDEV, underperform on average whereas the equal risk approaches, EN, EVA and RP, seem to outperform the RANDOM portfolio.

We utilize the Fung-Hsieh factor model introduced in Fung and Hsieh (2001) to account for the systematic risk exposures of hypothetical portfolios that might drive the above results. Table 5 reports mean and median values of Fung-Hsieh α and t-statistic of α and the percentage of positive αs for each portfolio methodology. The P-values are estimated using the bootstrap methodology. The superscript star indicates that the t-statistic of α of a given portfolio approach exceeds that of the RANDOM portfolio at 99 per cent confidence level. The subscript star shows that the t-statistic of α of a given portfolio approach is lower than that of the RANDOM portfolio at 99 per cent confidence level.

The minimum risk approaches, MV and MDEV, have mean t-statistics of α of around 1.59 which is lower than 2.26, the mean t-statistic of α of the RANDOM portfolio. The equal risk approaches, EN, EVA and RP,

Table 5: Mean and median statistics of Fung-Hsieh factor-based analysis for 1999–2014

Portfolio construction approach	*α (%)*	*T-statistic of α*	*Percentage of positive α (%)*
Panel A. Mean values			
RANDOM	6.61	2.256	99.29
EN	6.63	2.427	99.74
EVA	5.06	2.374	99.20
RP	5.24	2.343	99.13
MV	3.16	1.583	91.80
MDEV	3.16	1.588	91.89
Panel B. Median values			
RANDOM	6.56	2.276	99.29
EN	6.57	2.444	99.74
EVA	4.90	2.398	99.20
RP	5.13	2.358	99.13
MV	2.89	1.567	91.80
MDEV	2.89	1.570	91.89

This table presents results of regressions of the out-of-sample returns with respect to the Fung-Hsieh five factor model. Performance measures include annualized α, t-statistic of α, percentage of positive α. Panel A reports mean values, Panel B displayes median values.

yield values between 2.34 and 2.43 that dominate the RANDOM portfolio. Median values in Panel B demonstrate similar results. The P-values estimated using the bootstrap methodology suggest that equal risk approaches dominate RANDOM portfolios and the minimum risk approaches are inferior to RANDOM portfolios based on the

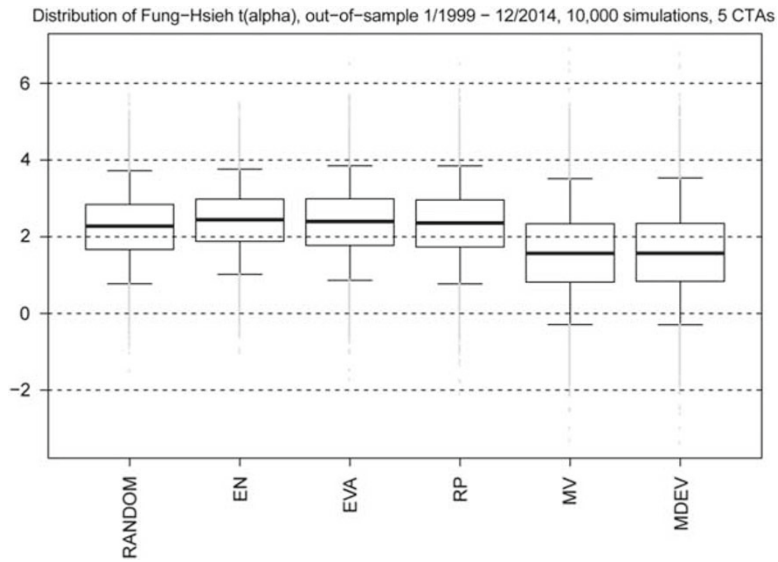

Figure 4: This figure shows distributions of the Fung-Hsieh (2001) five-factor *t*-statistic of α, generated using the large-scale simulation framework for the out-of-sample period between January 1999 and December 2014.

Fung-Hsieh *t*-statistic of α at the 99 per cent confidence level.

Figure 4 shows the distributions of the Fung-Hsieh *t*-statistic of α for each portfolio methodology.

The minimum risk approaches have heavy left tails and underperform the other methodologies on average. Therefore, the three key metrics of risk-adjusted performance, whether Sharpe, Calmar or the Fung-Hsieh *t*-statistic of α, suggest that the minimum risk portfolios are inferior and the equal risk approaches outperform the RANDOM portfolio on average.

Analysis of the marginal performance contribution of CTA portfolios to the investor's original portfolio

In this section, we evaluate the marginal impact of an investment in CTA portfolios for investors who hold a benchmark 60–40 portfolio of stocks and bonds. The comparison is done using Sharpe and Calmar ratios calculated for blended portfolios against the investor's original portfolio. First, we consider marginal contribution by comparing the marginal change in performance of a 90–10 blended portfolio that replaces 10 per cent of the original portfolio allocation with the CTA portfolios from the simulation using Sharpe and Calmar ratios. Then, we investigate the impact of the allocation to the CTA portfolios on the performance of the blended portfolios.

Relative performance of a 90–10 blended portfolio

Table 6 reports the average Sharpe and Calmar ratios of the blended portfolios and the percentage of simulations of blended portfolios that result in Sharpe and Calmar ratios that are superior to those of the original 60–40 portfolio.

The robustness of portfolio benefits stemming from an investment in CTAs is striking. Blended portfolios have higher Sharpe and Calmar ratios in at least 89 per cent of the scenarios among the worst performing minimum risk portfolios. Equal-risk portfolios have higher Sharpe and Calmar ratios in over 97 per cent of scenarios, and the improvement in average Sharpe ratios is as high as 10 per cent, with the original Sharpe improving from 0.376 to 0.41. Similarly, the equal-risk methodologies

Table 6: Portfolio contribution of CTA investments to the original investor portfolio 1999–2014

Portfolio construction approach	*Sharpe*	*Calmar*	*Improvement in Sharpe (%)*	*Improvement in Calmar (%)*
Panel A. Mean values				
RANDOM	0.416	0.108	96.58	98.46
EN	0.416	0.108	98.60	99.30
EVA	0.409	0.104	97.60	98.28
RP	0.410	0.105	97.53	98.14
MV	0.399	0.100	89.23	89.64
MDEV	0.399	0.100	89.14	89.46
Panel B. Median values				
RANDOM	0.415	0.107	96.58	98.46
EN	0.416	0.107	98.60	99.30
EVA	0.408	0.103	97.60	98.28
RP	0.410	0.104	97.53	98.14
MV	0.397	0.099	89.23	89.64
MDEV	0.397	0.099	89.14	89.46

This table reports results of marginal contribution analysis. The original investor portfolio is represented by 60–40 portfolio of stocks and bonds. It has delivered Sharpe ratio of 0.376 and Calmar ratio of 0.092 over the period 1999–2014. The first column presents Sharpe ratio of a blended portfolio that replaces 10 per cent allocation of the original portfolio with 10 per cent of the CTA portfolios constructed in the simulation framework. The second column reports Calmar ratio of blended portfolios. The third and fourth columns reports the percentage of time blended portfolios have higher Sharpe and Calmar ratios than then those of the original portfolio. Panel A reports mean values, Panel B displayes median values.

improve the average Calmar ratio by 10 per cent from 0.092 to over 0.1. Interestingly, a naïve diversification EN approach performs slightly better in terms of marginal performance contribution even though it marginally underperforms as a standalone investment. MBB perform analysis by market environment that can potentially give additional insight into the robustness of performance across market regimes. For brevity it is excluded here.[19]

Analysis of marginal performance contribution is important, particularly when an investor already has exposure to a large number of systematic sources of return in his or her well-diversified portfolio. In that situation, strategies that harvest the same sources of return can look very attractive as standalone investments but do not improve the risk-adjusted return of the investor's portfolio. The framework employed here is flexible and can utilize an investor's existing portfolio as a benchmark against which the marginal contribution of hedge fund portfolios can be measured.

The impact of the size of the allocation to CTA portfolios on the performance of blended portfolios

By evaluating the impact of allocation weights on performance, the framework can be used to optimally allocate to hedge fund portfolios given an investor's specific preferences and constraints. This study considers the performance of blended portfolios that have allocations between 5 per cent and 60 per cent to CTA investments. Table 7 reports the performance of blended portfolios stated in terms of Sharpe ratio. Panel A reports the percentage of simulations that improves the Sharpe ratio over the original 60–40 portfolio of stocks and bonds. Panel B reports mean Sharpe ratios and Panel C reports median Sharpe ratios of the blended portfolios.

Average Sharpe ratios increase until the allocation to CTA portfolios reaches 40–50 per cent and declines thereafter. However, the improvement that comes with a higher allocation to CTA portfolios also comes with a higher risk. While a MV portfolio improves the Sharpe ratio of the investor portfolio in 89.6 per cent of scenarios with a 5 per cent allocation to CTA portfolios, that number declines to 74 per cent at a 60 per cent allocation level. Similarly, the percentage of positive contribution scenarios declines from 98.7 to 81.6 per cent for the EN approach as the allocation to CTA investments grows from 5 to 60 per cent. Figure 5 shows the distribution of the out-of-sample Sharpe ratios of the blended portfolios.

It is important to note that the framework implicitly assumes that the performance of the investor's original portfolio can be expressed by a single time series or a single outcome, completely ignoring the role of luck because of active management decisions in the investor's portfolio.[20] A joint simulation of the investor's portfolio management

Table 7: Sharpe ratios of blended portfolios

	Allocation to CTA portfolios						
Portfolio construction approach	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*	*60%*
Panel A. Percentage of scenarios with higher sharpe							
RANDOM	97.0	96.6	95.2	92.9	88.6	82.3	73.4
EN	98.7	98.6	97.9	96.6	93.9	89.1	81.6
EVA	97.9	97.6	97.1	96.1	94.5	91.6	86.8
RP	97.7	97.5	96.8	95.6	93.5	90.4	84.9
MV	89.7	89.2	87.8	85.7	83.5	79.7	74.1
MDEV	89.6	89.1	87.9	85.9	83.4	79.7	74.0
Panel B. Mean							
RANDOM	0.396	0.416	0.454	0.482	0.494	0.487	0.463
EN	0.396	0.416	0.456	0.489	0.507	0.506	0.487
EVA	0.392	0.409	0.443	0.477	0.504	0.519	0.516
RP	0.393	0.410	0.447	0.482	0.509	0.522	0.517
MV	0.387	0.399	0.425	0.450	0.472	0.485	0.483
MDEV	0.387	0.399	0.424	0.449	0.470	0.483	0.480
Panel C. Median							
RANDOM	0.396	0.415	0.453	0.482	0.493	0.486	0.460
EN	0.396	0.416	0.456	0.488	0.506	0.504	0.483
EVA	0.391	0.408	0.442	0.475	0.502	0.517	0.513
RP	0.392	0.410	0.445	0.480	0.508	0.520	0.515
MV	0.386	0.397	0.420	0.444	0.464	0.477	0.475
MDEV	0.386	0.397	0.420	0.443	0.463	0.475	0.474

This table reports performance of blended portfolios for 1999–2014. Panel A reports percentage of scenarios with Sharpe ratio that exceeds Sharpe ratio of the investor's original portfolio. Panel B report cross-sectional mean of Sharpe ratios of blended portfolios. Panel C reports cross-sectional median of Sharpe ratios of blended portfolios.

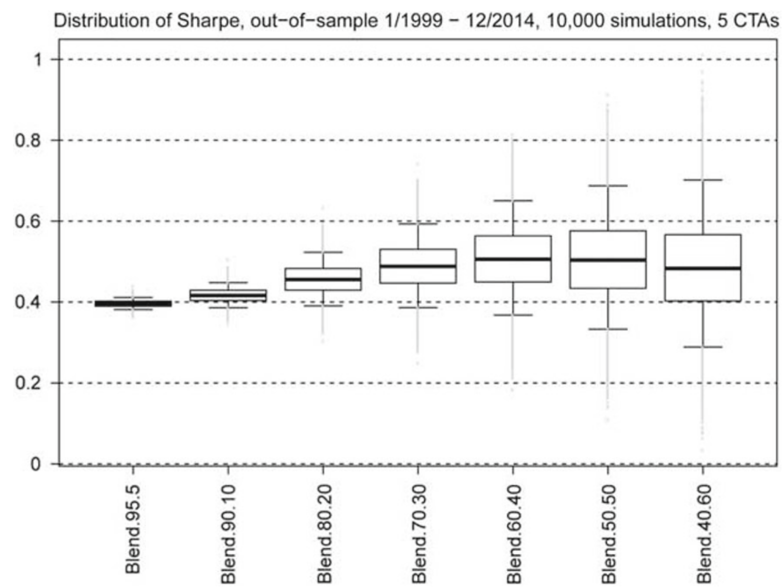

Figure 5: This figure shows distributions of the Sharpe ratios of blended portfolios of the original investor's portfolio of stocks and bonds and the hypothetical portfolios, generated using the large-scale simulation framework for the out-of-sample period between January 1999 and December 2014.

techniques applied to the original portfolio constituents and the hedge fund portfolios has the potential to better account for luck in both types of investments but requires additional assumptions that are outside the scope of this article.

Table 8: Calmar ratios of blended portfolios

	Allocation to CTA portfolios						
Portfolio construction approach	*5%*	*10%*	*20%*	*30%*	*40%*	*50%*	*60%*
Panel A. Percentage of scenarios with higher Sharpe							
RANDOM	98.5	98.5	98.0	97.0	95.9	94.2	91.6
EN	99.3	99.3	99.1	98.8	98.1	97.1	95.7
EVA	98.3	98.3	98.2	98.0	97.6	96.7	95.5
RP	98.1	98.1	98.0	97.7	97.0	96.1	94.8
MV	89.6	89.6	89.3	88.1	86.2	84.0	81.0
MDEV	89.5	89.5	89.1	87.9	86.2	83.9	81.3
Panel B. Mean							
RANDOM	0.099	0.108	0.128	0.153	0.182	0.211	0.224
EN	0.099	0.108	0.129	0.154	0.186	0.220	0.240
EVA	0.097	0.104	0.120	0.139	0.164	0.195	0.225
RP	0.098	0.105	0.121	0.142	0.167	0.199	0.228
MV	0.095	0.100	0.110	0.122	0.136	0.153	0.172
MDEV	0.095	0.100	0.110	0.122	0.136	0.152	0.171
Panel C. Median							
RANDOM	0.099	0.107	0.127	0.150	0.176	0.199	0.207
EN	0.099	0.107	0.128	0.152	0.181	0.210	0.225
EVA	0.097	0.103	0.119	0.137	0.160	0.186	0.212
RP	0.098	0.104	0.121	0.140	0.163	0.190	0.214
MV	0.095	0.099	0.109	0.119	0.130	0.142	0.154
MDEV	0.095	0.099	0.109	0.119	0.130	0.142	0.154

This table reports performance of blended portfolios for 1999–2014. Panel A reports percentage of scenarios with Calmar ratio that exceeds Calmar ratio of the investor's original portfolio. Panel B report cross-sectional mean of Calmar ratios of blended portfolios. Panel C reports cross-sectional median of Calmar ratios of blended portfolios.

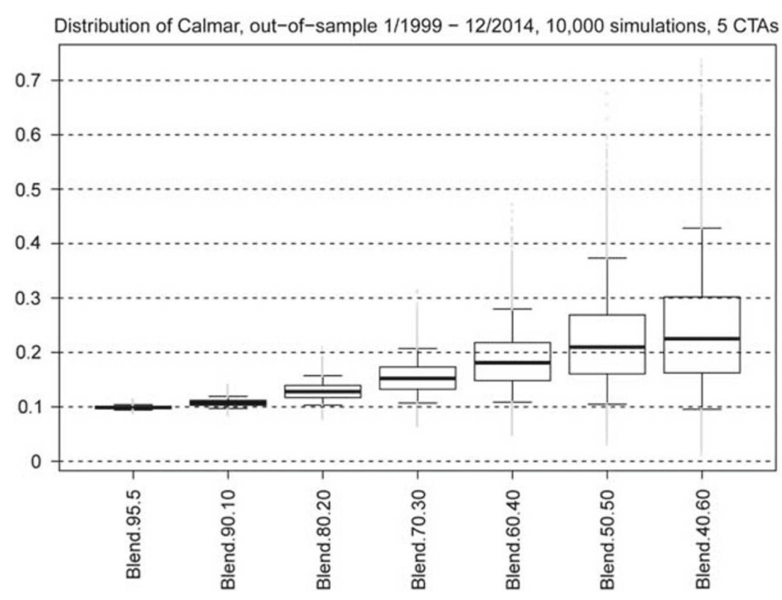

Figure 6: This figure shows distributions of the Calmar ratios of blended portfolios of the original investor's portfolio of stocks and bonds and the hypothetical portfolios, generated using the large-scale simulation framework for the out-of-sample period between January 1999 and December 2014.

Table 8 reports the performance of the blended portfolios stated in terms of Calmar ratio.

Panel A reports the percentage of simulations that improve the Calmar ratio over the original 60–40 portfolio of stocks and bonds. Panel B reports the mean Calmar ratios and Panel C reports the median Calmar ratios of the blended portfolios.

The average Calmar ratio grows monotonically with additional allocation to CTA investments without reaching an intermediate peak as in the case of Sharpe ratios. However, the improvement comes with higher risk as indicated by declining percentages scenarios with superior Calmar ratios. Figure 6 shows the distribution of the out-of-sample Calmar ratios of the blended portfolios.

The optimal allocation choice depends on the specific preferences of individual investors and their aversion to risk. Investors who value average performance will tend to pay more attention to the means and medians of the performance distributions of the blended portfolios. By contrast, investors who are very risk averse will put more weight on the characteristics of the left tails.

CONCLUDING REMARKS

This article introduces a quantitative large-scale simulation framework for the robust and reliable evaluation of hedge fund investments with real life constraints by institutional investors. This methodology is implementable and incorporates common investment constraints when creating and rebalancing portfolios. The framework is customizable to the preferences and constraints of individual investors, investment objectives, rebalancing periods and the desired number of funds in a portfolio and can include a large number of portfolio construction approaches. Thus, the methodology can benefit portfolio managers, investment officers, board members and consultants who make hedge fund investment decisions.

As an illustration of the framework, we applied it to a subset of hedge funds in managed futures revealing a strikingly significant portfolio contribution of CTA investments to a typical 60–40 portfolio of stocks and bonds over the period from 1999 to 2014, though this contribution is much less significant during the exceptional period between 2009 and 2014, when the benchmark portfolio delivered a Sharpe ratio of 1.12. This finding is robust across a large set of parameters and all portfolio construction methodologies considered in the study. The empirical results suggest that equal-risk portfolios of CTAs outperform minimum risk approaches out-of-sample whether as standalone investments or as diversifiers to the investor's benchmark portfolio.

While the empirical findings can immediately benefit institutional investors who seek to enhance performance through better diversification and portfolio construction, this analysis is merely one application of the flexible large-scale simulation methodology that can be utilized more broadly to examine a large number of portfolio management techniques subject to real life constraints.

ACKNOWLEDGEMENTS

The authors are grateful to the Investment team at Efficient and two anonymous referees for helpful comments. The authors appreciate the research assistance and valuable feedback of Kevin Townsend.

NOTES

1. Hedge funds and CTAs report their performance monthly and it takes several weeks to finalize end-of-month performance values. This issue is similar to the delay in reporting of the accounting book value explicitly addressed in Fama and French (1992). Studies by Jaganathan *et al* (2010), Agarwal and Naik (2000), Kosowksi *et al* (2007) are a few examples of papers that fail to account for the hedge fund reporting delay.
2. Jaganathan *et al* (2010) consider funds with any amount of AUM and track trecile portfolios with as many as 252 funds.
3. Fund selection criteria can incorporate performance-based ranking as in Molyboga *et al* (2015).
4. The framework is flexible and can incorporate customized performance measures selected by the investor. While the Fung-Hsieh (2001) five factor model is relevant for managed futures, the Fung-Hsieh eight factor model can be more appropriate for other types of hedge funds. MBB

evaluate performance using second order stochastic dominance which is particularly relevant because investors are often unaware of their own utility functions as reported in Elton and Gruber (1987). Levy and Sarnat (1970) and Fischmar and Peters (2006) suggest using stochastic dominance as an alternative to mean-variance analysis.

5. See Chekhlov *et al* (2005) for a formal definition of the maximum drawdown. It is typically defined as the largest peak-to-valley loss and represents a risk measure that is commonly used by practitioners. Calmar ratio is defined as the ratio of annualized excess return to the maximum drawdown.
6. Though in this paper marginal portfolio contribution is measured using Sharpe and Calmar ratios, in general it should be evaluated relative to the specific investment objectives of the investor. For example, a university endowment may target returns that exceed the university's spending rate over a market cycle. The framework can incorporate investor-specific performance metrics of marginal portfolio contribution.
7. According to the BarclayHedge Group which monitors assets under management, CTAs were managing \$310 million in 1980, \$10.5 billion in 1990 and \$330 billion in the first quarter of 2015.
8. The results of the sub-sample analysis are available upon request.
9. For details, see Appendix A: Data cleaning.
10. Typically funds go through an incubation period during which they build a track record using proprietary capital. Fund managers choose to start reporting to a CTA database to raise capital from outside investors only if the track record is attractive and they are allowed to 'backfill' the returns generated before their inclusion in the database. Since funds with poor performance are unlikely to report returns to the database, incubation/backfill bias results.
11. See Appendix B for technical definitions of the risk-based approaches.
12. See MBB for a detailed description of the hedge fund reporting delay.
13. The framework is flexible – the number of funds in a portfolio, rebalancing frequency, AUM threshold levels and other parameters can be customized to reflect each investor's preferences and constraints.
14. Calmar ratio is defined as the ratio of the annualized excess return to the maximum historical drawdown.
15. The framework is flexible in comparing any two approaches to each other but requires performing additional bootstrapping simulations based on an investor's particular areas of interest.
16. Jensen's inequality states that $Eg(X) \leqslant g(Ex)$ for any concave function g such as the Sharpe ratio. See Rudin (1986) for a detailed explanation of the Jensen's inequality.
17. See MBB for detailed examples of employing first and second order stochastic dominance to evaluate distributions of out-of-sample performance within a large-scale simulation framework.
18. The *P*-value is estimated by calculating the percentage of bootstrapped simulations of RANDOM portfolios that outperform the other portfolio methodologies for a given performance metric. For example, the *P*-value of 16 for EN in the Return category suggests that 16 per cent of bootstrapped simulations have a mean return that is higher than that of EN. Therefore, we fail to reject the hypothesis of RANDOM portfolios having a mean return that is lower than that of EN. That intuitively makes sense because random portfolios should have the same return as equal portfolios on average. We compare RANDOM portfolios to bootstrapped RANDOM portfolios for robustness. The *P*-values indicate that we cannot reject the hypothesis that the RANDOM portfolio is better or worse than the bootstrap RANDOM portfolios at any reasonable confidence level.
19. We have performed a sub-sample analysis to evaluate the marginal contribution of a 10 per cent allocation to managed futures during periods of relatively poor and relatively good performance. A 60–40 portfolio produced a Sharpe ratio of −0.09 and a Calmar ratio of −0.025 during the relatively poor period between 1999 and 2008. During the exceptional period between 2009 and 2014, the benchmark portfolio delivered a Sharpe ratio of 1.12 and a Calmar ratio of 0.77. During the 10-year period between 1999 and 2008, all portfolio construction approaches would have added value as measured in terms of average Sharpe and Calmar ratios with the equal-risk portfolios producing the best results. During the 6-year period between 2009 and 2014, the blended portfolios have approximately the same average Sharpe ratios and slightly better Calmar ratios than the benchmark portfolio with the minimum risk approaches delivering the best results. The sensitivity of relative results to an evaluation window is common. For example, Anderson *et al* (2012) compare four investment strategies and find that the specific start and end dates of a backtest can have a material impact on the results. This study introduces a methodology that can be used for the evaluation of portfolio construction approaches with real-life constraints and is strengthened by the sub-sample analysis. Results of the sub-sample analysis are available upon request.
20. Since we evaluate the role of luck in active management decisions, we consider that a passive 60–40 portfolio of stocks and bonds that utilizes the S&P 500 Total Return index and the JPM Global Government Bond Index has no luck associated with it.

REFERENCES

Abrams, R., Bhaduri, R. and Flores, E. (2009) Lintner Revisited: A Quantitative Analysis of Managed Futures in an Institutional Portfolio. CME Group Working Paper.

Ackermann, C., McEnally, R. and Ravenscraft, D. (1999) The performance of hedge funds: Risk, return, and incentives. *Journal of Finance* 54(3): 833–874.

Agarwal, V. and Naik, N. (2000) Multiperiod performance persistence analysis of hedge funds. *Journal of Financial and Quantitative Analysis* 35(3): 327–342.

Anderson, R., Bianchi, S. and Goldberg, L. (2012) Will my risk parity strategy outperform? *Financial Analysts Journal* 68(6): 75–93.

Anson, M. (2011) The evolution of equity mandates in institutional portfolios. *Journal of Portfolio Management* 37(4): 127–137.

Chekhlov, A., Uryasev, S. and Zabarankin, M. (2005) Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance* 8(1): 13–58.
Chen, P., O'Neill, C. and Zhu, K. (2005) Managed Futures and Asset Allocation. Ibbotson Working Paper.
Clarke, R., de Silva, H. and Thorley, S. (2013) Risk parity, maximum diversification, and minimum variance: An analytic perspective. *Journal of Portfolio Management* 39(3): 39–53.
DeMiquel, V., Garlappi, L. and Uppal, R. (2009) Optimal versus naïve diversification: How efficient is the $1/N$ portfolio strategy. *Review of Financial Studies* 22(5): 1915–1953.
Efron, B. (1979) Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7(1): 1–26.
Efron, B. and Gong, G. (1983) A leisurely look at the bootstrap, the jackknife, and cross validation. *The American Statistician* 37(1): 36–48.
Elton, E. and Gruber, M. (1987) *Modern Portfolio Theory and Investment Analysis*. 3rd edn. New York: Wiley.
Fama, E. and French, K. (1992) The cross-section of expected stock returns. *Journal of Finance* 47(2): 427–465.
Fama, E. and French, K. (2010) Luck versus skill in the cross-section of mutual fund returns. *Journal of Finance* 65(5): 1915–1947.
Fischmar, D. and Peters, C. (2006) Portfolio analysis of stocks, bonds, and managed futures using compromise stochastic dominance. *Journal of Futures Markets* 11(3): 259–270.
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.
Hallerbach, W. (2012) A proof of optimality of volatility weighting over time. *Journal of Investment Strategies* 1(4): 87–99.
Jagannathan, R. and Ma, T. (2003) Risk reduction in large portfolios: why imposing the wrong constraints helps. *Journal of Finance* 58(4): 1651–1683.
Jagannathan, R., Malakhov, A. and Novikov, D. (2010) Do hot hands exist among hedge fund managers? An empirical evaluation. *Journal of Finance* 65(1): 217–255.
Joenvaara, J., Kosowski, R. and Tolonen, P. (2012) New 'Stylized Facts' About Hedge Funds and Database Selection Bias, Working Paper.
Kat, H. (2004) Managed futures and hedge funds: A match made in heaven. *Journal of Investment Management* 2(1): 1–9.
Kosowski, R., Naik, N. and Teo, M. (2007) Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *Journal of Financial Economics* 84(1): 229–264.
Kritzman, M., Page, S. and Turkington, D. (2010) In defense of optimization: The fallacy of $1/N$. *Financial Analysts Journal* 66(2): 31–3.
Levy, H. and Sarnat, M. (1970) Alternative efficiency criteria: An empirical analysis. *Journal of Finance* 25: 1153–1158.
Lintner, J. (1996) The potential role of managed commodity-financial futures accounts (and/or funds) in portfolios of stocks and bonds. In: C.C. Peters and B. Warwick (eds.) *The Handbook of Managed Futures: Performance, Evaluation and Analysis.*, New York: McGraw-Hill Professional, pp. 99–137.
Maillard, S., Roncalli, T. and Teiletche, J. (2010) The properties of equally weighted risk contribution portfolios. *Journal of Portfolio Management* 36(4): 60–70.
Molyboga, M., Baek, S. and Bilson, J. (2015) Assessing Hedge Fund Performance with Institutional Constraints, Working Paper.
Opdyke, J. (2007) Comparing sharpe ratios: So where are the P-values? *Journal of Asset Management* 8(5): 308–336.
Qian, E. (2013) Are risk-parity managers at risk parity? *Journal of Portfolio Management* 40(1): 20–26.
Rudin, W. (1986) *Real and Complex Analysis (Higher Mathematics Series)*. 3rd edn. New York: McGraw-Hill.

APPENDIX A

Data cleaning

After excluding all funds from the BarclayHedge database that are multi-advisors or benchmarks, we select only those funds that report returns net of all fees for the period between December 1993 and December 2014. Our study considers 4673 funds with 1013 active and 3660 defunct funds. We performed a few additional data filtering procedures to improve data quality and make the results practical for institutional investors. First, we eliminated null returns at the end of the track records of defunct fund. Then we excluded managers with less than 24 months of data which limited the data set to 3223 funds. In addition, we eliminated all funds with maximum AUM of less than $10 million which further limited the data set to 1937 funds. Finally, we excluded funds with one or more monthly return in excess of 100 per cent that resulted in the final pool of 1927 funds of which 604 were live and 1323 were defunct.

APPENDIX B

Risk-based allocation approaches

In this study we consider three equal-risk and two minimum risk approaches. They include EN, EVA, classic RP, MV and MDEV methodologies.

1. EN allocation is a simple equal weight (or naïve diversification) approach:

$$w_i = \frac{1}{N}$$

where N is the number of funds in the portfolio and w_i is the weight of fund i.

2. EVA allocation is similar to the EN approach but exposure to each fund is adjusted for the fund's volatility which is estimated using the standard deviation of its in-sample excess returns:

$$w_i = \frac{\frac{1}{\sigma_i}}{\sum_{j=1}^{N}\left[\frac{1}{\sigma_j}\right]}$$

3. Classic RP is the solution to the following optimization problem:

$$\min_w \sum_{i=1}^{N}\left(\frac{\partial\sigma}{\partial w_i}\frac{w_i}{\sigma} - \frac{1}{N}\right)^2$$

such that $\sum_{i=1}^{N} w_i = 1$, $w_i \geqslant 0$, and $\sigma = \sqrt{w'\Sigma w}$ represents portfolio volatility with Σ, the sample covariance matrix, calculated using the in-sample excess returns.

4. MV is the solution to the following optimization problem:

$$\mathrm{Min}_w \sigma$$

such that $\sum_{i=1}^{N} w_i = 1$, $w_i \geqslant 0$.

5. MDEV is the solution to the following optimization problem:

$$\mathrm{Min}_w \sigma_T$$

such that $\sum_{i=1}^{N} w_i = 1$, $w_i \geqslant 0$, where $\sigma_T = \sqrt{(1/(N-1))\sum_{j=1}^{N} x_j^2 I_{\{x_j<0\}}}$, and x_j are the fund's monthly returns during the N-month in-sample period with $j = 1,\ldots, N$.

6. Random portfolio (RANDOM) is used as a benchmark approach to portfolio allocation. First, a random number x_i between 0 and 1 is generated. Then random portfolio weights are normalized by setting $w_i = x_i / \sum_{j=1}^{N} x_j$.

This work is licensed under a Creative Commons Attribution 3.0 Unported License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/