



Management Science

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

Do Hedge Funds Outperform Stocks and Bonds?

Turan G. Bali, Stephen J. Brown, K. Ozgur Demirtas,

To cite this article:

Turan G. Bali, Stephen J. Brown, K. Ozgur Demirtas, (2013) Do Hedge Funds Outperform Stocks and Bonds?. Management Science 59(8):1887-1903. <http://dx.doi.org/10.1287/mnsc.1120.1689>

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

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

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

Copyright © 2013, INFORMS

Please scroll down for article—it is on subsequent pages



INFORMS is the largest professional society in the world for professionals in the fields of operations research, management science, and analytics.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Do Hedge Funds Outperform Stocks and Bonds?

Turan G. Bali

McDonough School of Business, Georgetown University, Washington, DC 20057, tgb27@georgetown.edu

Stephen J. Brown

Leonard N. Stern School of Business, New York University, New York, New York 10012; and
University of Melbourne, Parkville 3010, Victoria, Australia, sbrown@stern.nyu.edu

K. Ozgur Demirtas

Finance at the School of Management, Sabanci University, Orhanli, Tuzla 34956, Istanbul, Turkey,
ozgurdemirtas@sabanciuniv.edu

Hedge funds' extensive use of derivatives, short selling, and leverage and their dynamic trading strategies create significant nonnormalities in their return distributions. Hence, the traditional performance measures fail to provide an accurate characterization of the relative strength of hedge fund portfolios. This paper uses the utility-based nonparametric and parametric performance measures to determine which hedge fund strategies outperform the U.S. equity and/or bond markets. The results from the realized and simulated return distributions indicate that the long/short equity hedge and emerging markets hedge fund strategies outperform the U.S. equity market, and the long/short equity hedge, multistrategy, managed futures, and global macro hedge fund strategies dominate the U.S. Treasury market.

Key words: hedge funds; stocks; bonds; almost stochastic dominance; manipulation-proof performance measure

History: Received May 10, 2012; accepted November 16, 2012, by Wei Jiang, finance. Published online in

Articles in Advance March 4, 2013.

1. Introduction

Hedge funds employ a wide variety of dynamic trading strategies, and make extensive use of derivatives, short selling, and leverage (see Aragon and Martin 2012). Most hedge fund investment strategies aim to achieve a positive return on investment whether markets are rising or falling.¹ Hedge funds' frequent employment of derivatives, short selling, and leverage create significant skewness and kurtosis in their return distributions. Because of their speculative bets and active trading, the probability distribution of hedge fund returns is often characterized by asymmetric and leptokurtic behavior (fat tails) that are affected by event risks.² Hedge funds, for example, suffered large losses during the 1997 Asian currency crisis, the 1998 Russian debt crisis, and the recent credit crunch. As an asset class, hedge funds were

not alone in experiencing significant losses during the recent financial crises. However, their dynamic trading, use of derivatives, and arbitrage strategies generate payoff structures that are often nonlinear functions of the returns of the underlying assets, which in some cases magnify the extent of the losses that occur.

Scott and Horvath (1980) show that under weak assumptions with respect to investors' utility functions, investors prefer high first and third moments (mean and skewness) and low second and fourth moments (standard deviation and kurtosis). In other words, investors have preference for assets with higher expected return and larger positive skewness, whereas they dislike assets with higher volatility and higher kurtosis. A negative skew is often held to be a measure of left tail risk to the extent that it is consistent with a long left tail of the return distribution, with the bulk of the values (possibly including the median) lying to the right of the mean. High kurtosis means that more variance can be attributed to infrequent extreme returns and is consistent with a sharper peak and longer tails than would be implied by a Normal distribution. Large negative skewness and large values of kurtosis generate high downside risk particularly for hedge funds.

This paper investigates whether hedge fund strategies dominate the U.S. equity and bond markets. Because the empirical return distribution of hedge

¹ Despite the fact that hedge funds are marketed as absolute return or market-neutral investments that generate positive returns in both good and bad market conditions, Asness et al. (2001), Chen and Liang (2007), Patton (2009), and Bali et al. (2011) show that hedge fund returns are exposed to market factors.

² Mitchell and Pulvino (2001), Malkiel and Saha (2005), Bali et al. (2007), and Fung et al. (2008) provide evidence for departures from normality for the hedge funds' return distributions with significant tail risk. Brown et al. (2012) show that this tail risk exposure may not be diversifiable, which suggests that compensation for tail risk exposure could explain hedge fund returns.

fund portfolios as well as the distributions of equity and bond returns exhibit significant departures from normality, the mean-variance (MV), Sharpe ratio, and alpha comparisons do not provide an accurate characterization of the relative strength of different asset classes. It is well known that the return distributions of financial securities are skewed and fat tailed with significant excess kurtosis, which motivates us to consider alternative methodologies when comparing the performance of hedge fund portfolios with the U.S. equity index as well as the short-term and long-term Treasury securities.

This paper is the first to utilize the almost stochastic dominance (ASD) approach of Leshno and Levy (2002), and the manipulation-proof performance measure (MPPM) of Goetzmann et al. (2007) to examine the relative performance of hedge fund portfolios. The ASD approach does not require a parametric specification of investors' preferences and does not make any assumptions about asset returns. Instead, ASD imposes general conditions regarding preferences and considers the entire return distribution with high-order moments. In addition to the utility-based nonparametric approach of Leshno and Levy (2002), we use the MPPM of Goetzmann et al. (2007), which is a parametric approach but similar to the ASD approach is also utility based. We contribute to the literature by using these more generalized approaches, which robustly and accurately identify the relative performance of hedge fund strategies with respect to the equity and bond markets.

The results from the realized and simulated return distributions indicate that popular hedge fund strategies, long/short equity hedge and emerging markets, outperform the U.S. equity market. However, the remaining nine hedge fund indices considered in the paper do not generate superior performance over the S&P 500 index. For short- to medium-term investment horizons, the long/short equity hedge, multistrategy, managed futures, and global macro hedge fund strategies dominate the U.S. Treasury market. For long investment horizons, most hedge fund strategies perform better than the one-month and 10-year Treasury securities.

This paper is organized as follows. Section 2 describes the decision rules and investors' preferences. Section 3 evaluates the performance of hedge fund strategies based on the ASD approach. Section 4 presents results from the traditional and manipulation-proof performance measures. Section 5 provides results from downside risk-adjusted performance measures. Section 6 investigates the relative strength of hedge fund portfolios using time-varying conditional distributions. Section 7 concludes this paper.

2. Decision Rules

2.1. Classical Decision Rules

Classical investment decision-making rules may fail to provide a preference between two portfolios, although it is clear that all investors or almost all investors would choose one portfolio over the other. The reason for this failure can be demonstrated by a simple example with two portfolios, H and L :

$$H: \mu_H = 1,000\%, \quad \sigma_H = 10.1\%;$$

$$L: \mu_L = 1\%, \quad \sigma_L = 10\%,$$

where H and L represent portfolios with high and low expected returns, respectively. The expected returns on portfolio H and L are denoted by μ_H and μ_L , respectively, whereas the corresponding standard deviations are denoted by σ_H and σ_L . For portfolio H to dominate portfolio L by the mean-variance rule, both $\mu_H \geq \mu_L$ and $\sigma_H \leq \sigma_L$ must hold. As presented in the above example, portfolio H has a significantly higher expected return and slightly higher volatility, which implies no dominance in the MV framework. However, in a randomly selected group of investors, all would clearly choose portfolio H over portfolio L because the decline in their expected utility from the slightly higher volatility is much less than the increase in their expected utility from the significantly higher expected return. Hence, the MV rule is unable to distinguish between two investment choices though all investors would prefer H over L .

Further insight can be gained using cash flows for the H and L portfolios as shown in Table 1, where the probability of each state (low, medium, and high) and the corresponding cash flows of the H and L portfolios are given. For portfolio H to dominate portfolio L by the first-order stochastic dominance (FSD) rule, the cumulative distribution of portfolio H should strictly plot below the cumulative distribution of portfolio L . As shown above, for most of the probable region, the cumulative distribution of H plots below that of L . However, there is a small violation area in the low state (in which portfolio H earns \$1, whereas portfolio L earns \$2). Therefore, portfolio H does not dominate portfolio L by the FSD rule. Yet, in any randomly selected sample of investors, one should not be surprised to find that all investors would choose H .

Table 1 Cash Flows for the H and L Portfolios

State	Probability (%)	L (\$)	H (\$)
Low	1	2	1
Medium	2	3	3
High	97	4	1 million

Note. This table presents the probabilities and cash flows for the H and L portfolios for three states (low, medium, high).

over L . On the other hand, for an investor with a utility function $u(x) = x$ for $x < 2$ and $u(x) = 2$ for $x \geq 2$, H does not necessarily dominate L . But, this utility function implies that the investor is indifferent between getting \$2 and \$1 million, hence, it is called a *pathological* utility function, which is economically irrelevant.

The above examples present the limitations of the decision rules based on extreme representations of the paradoxes. However, these examples are not unique, as many other cases with such a paradox can easily be created. To solve this issue, Leshno and Levy (2002) proposed rules known as the ASD rules, which are appropriate after eliminating the pathological preferences. Using these ASD rules, one can avoid paradoxes demonstrated above. In this paper, we utilize the ASD rules as well as the classical investment decision rules to explain the phenomena of investing in hedge funds in an expected utility paradigm.

First-Order Stochastic Dominance. Let there be two risky portfolios, H and L , and denote F_H and F_L as the cumulative distribution of H and L , respectively. Portfolio H dominates portfolio L by FSD ($H \succeq_1 L$) if $F_H(r) \leq F_L(r)$ for all return values r and a strict inequality holds for at least some r . In other words, $H \succeq_1 L$ iff $E_H u(r) \geq E_L u(r)$ for all $u \in U_1$, where U_1 is the set of all nondecreasing differentiable real-valued functions.

Second-Order Stochastic Dominance. High expected return portfolio (H) dominates low expected return portfolio (L) by SSD ($H \succeq_2 L$) if $\int_{-\infty}^r [F_L(s) - F_H(s)] ds \geq 0$ for all r and a strict inequality holds for at least some r . In other words, $H \succeq_2 L$ iff $E_H u(r) \geq E_L u(r)$ for all $u \in U_2$, where U_2 is the set of all nondecreasing real-valued functions such that $u'' \leq 0$.

2.2. Almost Stochastic Dominance Rules

To introduce the concept of almost stochastic dominance rules, we first define the violation area. When considering whether H dominates L , the region where the cumulative distribution of H is above the cumulative distribution of L is called the violation area (denoted by V in Figure 1), which is the reason why H fails to dominate L .

More specifically, the violation area V is $\int_{r_1}^{r_2} [F_H(s) - F_L(s)] ds$,³ where the FSD violation range is given by

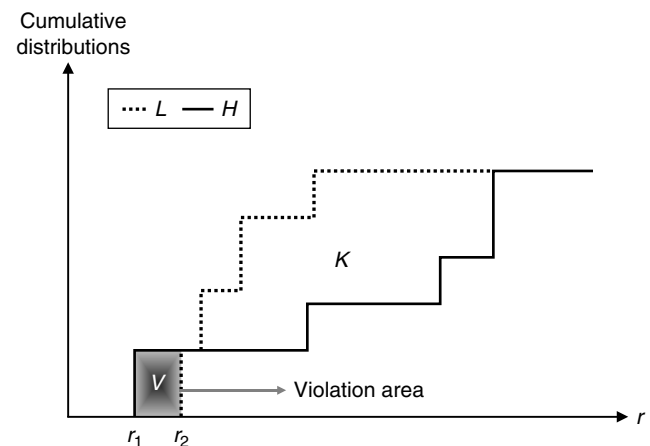
$$R_1(F_H, F_L) = \{s \in (r_1, r_2) : F_L(s) < F_H(s)\}. \quad (1)$$

We define the *empirical* violation area as

$$\varepsilon_1 = \frac{\int_{R_1} [F_H(s) - F_L(s)] ds}{\int_{\min}^{\max} |F_H(s) - F_L(s)| ds}, \quad (2)$$

³ Because of the violation area denoted by V , H does not dominate L by FSD, SSD, or MV rules because there are some utility functions that give a large weight to area V and a very small or zero weight to area K in Figure 1.

Figure 1 Almost First-Order Stochastic Dominance



Notes. Consider two cumulative distributions of the high- and low-return portfolios, H and L . Because of the violation area denoted by V , H does not dominate L by FSD, SSD, or MV rules. There are some utility functions that give a large weight to area V and a very small or zero weight to area K , where H is below L . Because almost all investors would prefer portfolio H , the AFSD rule is introduced, which reveals a dominance of H over L despite the fact that area V violates FSD.

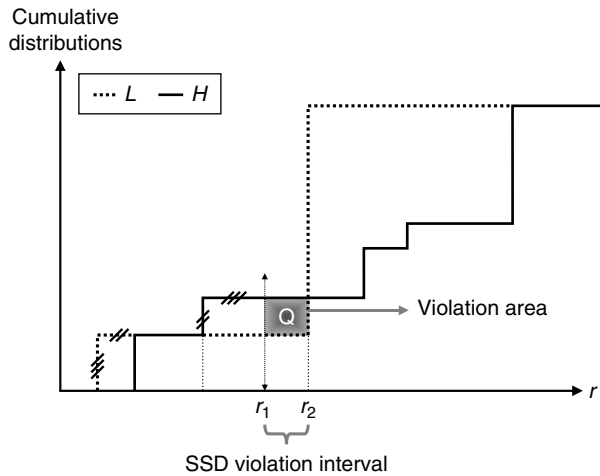
where F_H and F_L have a finite support $[\min, \max]$. In Equation (2), ε_1 is defined as the area above $[r_1, r_2]$ (area V in Figure 1) divided by the total *absolute* area enclosed between F_H and F_L (area $V + K$ in Figure 1).

Clearly $\varepsilon_1 = 0$ implies FSD. On the other hand, when $\varepsilon_1 > 0$, there is no FSD, but there may be almost FSD (AFSD). Whether H dominates L by AFSD depends on whether nonzero empirical violation area ε_1 is small enough. What is small enough is an empirical question. Suppose that for each investor i we observe the highest value of ε_1 (denoted by $\varepsilon_{1,i}^*$) such that this investor prefers H over L . Then, the minimum of $\varepsilon_{1,i}^*$ across all investors provides the critical value $\varepsilon_1^* = \min(\varepsilon_{1,i}^*)$. In other words, even when H is not preferred over L because of a small violation area, as long as this violation area estimated by the empirical data is smaller than the critical value (i.e., $\varepsilon_1 < \varepsilon_1^*$), we conclude that H is preferred over L by AFSD. As originally discussed by Levy et al. (2010), ε_1^* is obtained from a series of experimental studies and found to be 5.9% for the AFSD rule. More formally, AFSD is defined as follows.

Almost First-Order Stochastic Dominance. Define the set of preferences U_1^* as

$$U_1^*(\varepsilon_1) = \left\{ u \in U_1; u'(s) \leq \inf\{u'(s)\} \left[\frac{1}{\varepsilon_1} - 1 \right], \right. \\ \left. \forall s \in (\min, \max) \right\}, \quad (3)$$

such that U_1^* denotes the set of all nondecreasing utility functions excluding the pathological preferences. Then for all $u \in U_1^*$, H dominates L , iff $\varepsilon_1 \leq \varepsilon_1^*$. Note

Figure 2 Almost Second-Order Stochastic Dominance

Notes. This figure illustrates a case where H has a much higher mean than L , but because of the negative area within the interval $[r_1, r_2]$ (i.e., area Q), there is no SSD of H over L . Yet, if area Q is “relatively small” and $E_H(r) - E_L(r)$ is “relatively large,” ASSD may exist.

that this condition holds if and only if $E_H u(r) \geq E_L u(r)$ for all $u \in U_1^*$.

Figure 2 presents a case where H has a much higher mean than L , but because of the negative area within the return interval $[r_1, r_2]$ (i.e., area Q), there is no SSD of H over L because the integrated area between F_H and F_L becomes positive at the beginning of area Q (i.e., by the return value r_1).

Suppose that, in most of the return interval, F_H is below F_L as demonstrated in Figure 2. We define the SSD violation range by

$$R_2(F_H, F_L) = \left\{ s \in R_1(F_H, F_L) : \int_0^s [F_L(s) - F_H(s)] ds < 0 \right\}. \quad (4)$$

Then, the empirical SSD violation area, ε_2 , is defined as

$$\varepsilon_2 = \frac{\int_{R_2} [F_H(s) - F_L(s)] ds}{\int_{\min}^{\max} |F_H(s) - F_L(s)| ds}, \quad (5)$$

where ε_2 is defined as the area above $[r_1, r_2]$ (area Q in Figure 2) divided by the total absolute area enclosed between F_H and F_L . Similar to the critical ε_1^* value for the AFSD, ε_2^* is obtained from the minimum values of $\varepsilon_{2,i}^*$ across all investors. As described by Bali et al. (2009) and Levy et al. (2010), ε_2^* is obtained similarly from a series of experimental studies and found to be 3.2% for the almost second-order stochastic dominance (ASSD) rule.⁴ More formally, ASSD is defined as follows.

⁴ In the appendix, we conduct several simulation analyses to show that the critical values used in our empirical analysis are robust indicators of almost all investors' choice of a high-return portfolio over a low-return portfolio. Specifically, randomization and

Almost Second-Order Stochastic Dominance. Define the set of preferences U_2^* as

$$U_2^*(\varepsilon_2) = \left\{ u \in U_2; -u''(s) \leq \inf\{-u''(s)\} \left[\frac{1}{\varepsilon_2} - 1 \right], \right. \\ \left. \forall s \in (\min, \max) \right\}. \quad (6)$$

Then for all $u \in U_2^*$, H dominates L by ASSD, iff $\varepsilon_2 \leq \varepsilon_2^*$. Clearly if $\varepsilon_2 = 0$ then SSD is obtained. For $\varepsilon_2 > 0$, there is no SSD, but ASSD may be obtained.

3. Empirical Results

3.1. Data

The hedge fund data set is obtained from Lipper TASS database, and as of December 2011, it contains information on a total of 17,383 defunct and live hedge funds with total assets under management close to \$1.3 trillion. Between January 1994 and December 2011, of the 17,383 hedge funds that reported monthly returns to TASS, we have 10,587 funds in the defunct/graveyard database and 6,794 funds in the live hedge fund database. One interesting observation is the large size disparity seen among hedge funds, where the size of a fund is measured as the average monthly assets under management over the life of the fund. Based on our data, while the mean hedge fund size is \$118.3 million, the median hedge fund size is only \$31.8 million. This suggests the existence of very few hedge funds with very large assets under management, which again reflects the true hedge fund industry standards.⁵

Earlier studies (e.g., Brown et al. 1992, 1999; Ackermann et al. 1999; Liang 1999; Agarwal and Naik 2000; Fung and Hsieh 2004; Kosowski et al. 2007; Fung et al. 2008) recognize the sample selection bias issues inherent in all hedge fund research and address these issues in various ways. We follow this literature by including both live (6,794 funds) and dead funds (10,587) in our sample to eliminate survivorship bias. We find that if the returns of nonsurviving hedge funds (graveyard database) had not been included in the analyses, there would have been a survivorship bias of approximately 2% in average annual hedge fund returns.⁶ In dealing with the backfill bias

bootstrapping methodologies are used to examine the p -values corresponding to the critical values (ε_1^* and ε_2^*). In the online appendix (available at <http://faculty.msb.edu/tgb27/>), we estimate the confidence intervals using a large scale of simulation analyses.

⁵ See Liang (2003) and Bali et al. (2007).

⁶ This finding is comparable to those of earlier studies of hedge funds. Liang (2000) reports an annual survivorship bias of 2.24%, Edwards and Caglayan (2001) report an annual survivorship bias of 1.85%, Bali et al. (2011) report an annual survivorship bias of 1.74%, and Bali et al. (2012) report an annual survivorship bias of 1.91%.

Table 2 Descriptive Statistics

Asset	Mean (%)	Median (%)	Stdev (%)	Skewness	Kurtosis	Min (%)	Max (%)	JB	p-value
Convertible arbitrage	0.62	0.87	2.17	−3.36	29.71	−17.46	8.41	6,828.25	0.0000
Dedicated short bias	0.26	−0.06	4.09	0.83	6.49	−9.69	22.09	134.57	0.0000
Emerging markets	0.99	1.47	4.27	−0.95	6.92	−21.97	14.30	170.67	0.0000
Equity market neutral	0.71	0.71	0.91	−1.47	9.11	−4.54	2.54	413.72	0.0000
Event driven	0.81	1.18	1.64	−1.76	8.92	−7.65	4.21	427.38	0.0000
Fixed income arbitrage	0.68	0.85	1.13	−2.65	16.78	−7.15	2.83	1,961.06	0.0000
Fund of funds	0.49	0.56	1.55	−0.34	5.37	−5.52	5.65	54.62	0.0000
Global macro	0.74	0.72	1.62	0.68	4.00	−3.72	6.32	25.69	0.0000
Long/short equity hedge	0.99	1.10	2.53	−0.01	4.66	−8.43	10.19	24.67	0.0000
Managed futures	0.77	0.62	2.62	0.21	2.57	−5.10	7.46	3.23	0.1985
Multistrategy	0.87	0.90	1.19	−0.80	6.01	−4.40	4.16	104.50	0.0000
S&P 500	0.56	1.12	4.53	−0.64	3.89	−16.94	10.77	21.71	0.0000
One-month T-bill	0.26	0.31	0.17	−0.24	1.56	0.00	0.56	20.81	0.0000
10-year T-bond	0.55	0.63	2.10	−0.04	4.10	−6.68	8.54	10.96	0.0042

Notes. This table presents the descriptive statistics of monthly returns on the hedge fund portfolios, S&P 500 index, one-month Treasury bill, and 10-year Treasury bond for the sample period January 1994–December 2011. The table reports the mean, median, standard deviation, skewness, kurtosis, minimum, maximum, and Jarque–Bera (JB) statistics along with the *p*-values. $JB = n[(S^2/6) + (K - 3)^2/24]$ is a formal statistic for testing whether the returns are normally distributed, where *n* denotes the number of observations, *S* is skewness, and *K* is kurtosis. The JB statistic is distributed as the chi-square with two degrees of freedom.

(e.g., Aggarwal and Jorion 2010), we find that there is a one-year gap between the first performance date and the date that the fund is added to the database. We discover that the average annual return of hedge funds during the first year of existence is, in fact, 1.8% higher than the average annual returns in subsequent years. To avoid backfill bias, we follow Fung and Hsieh (2000) and delete the first 12-month return histories of all individual hedge funds in our sample. Last, to address the multiperiod sampling bias and to obtain sensible measures of risk for funds, we require that all hedge funds in our study have at least 24 months of return history (see Kosowski et al. 2007) to mitigate the impact of multiperiod sampling bias.⁷

After all of these requirements, we have 11,973 surviving and defunct funds in our sample (6,812 dead funds and 5,161 live funds). We compute the equal-weighted average returns of funds for each of the 11 investment styles reported in the TASS database to generate hedge fund indices:⁸ convertible arbitrage, dedicated short bias, emerging markets, equity market neutral, event driven, fixed income arbitrage, fund of funds, global macro, long/short equity hedge, managed futures, and multistrategy. The performance of the U.S. equity market is measured by the S&P 500 index returns, and the perfor-

mance of the short-term and long-term U.S. Treasury securities is proxied by the one-month and 10-year Treasury returns, respectively.⁹

3.2. Classical Investment Decision Making

In this section, we compare the performance of 11 hedge fund strategies with the S&P 500 index, one-month Treasury bills (T-bills), and 10-year Treasury bonds (T-bonds) using the classical selection rules based on the mean-variance and stochastic dominance criteria. First, we assess the relative performance of hedge fund investment styles with reference to the S&P 500 index. Table 2 shows that the average monthly return on the S&P 500 index is 0.56% and the standard deviation is 4.53% per month for the sample period January 1994–December 2011. Nine out of 11 hedge fund strategies have higher average return and lower standard deviation. Specifically, the long/short equity hedge, emerging markets, multistrategy, event driven, managed futures, global macro, equity market neutral, fixed income arbitrage, and convertible arbitrage strategies have average monthly returns in the range of 0.99% to 0.62%, with the monthly standard deviations ranging from 0.91% to 4.27%. Hence, based on Markowitz's (1952) MV criterion, we conclude that the aforementioned nine hedge fund strategies dominate the U.S. equity market.

However, as presented in Table 2, there are significant departures from normality in the return distribution of hedge fund portfolios. Specifically, the

⁷ In this paper, we do not analyze the biases related to self-reporting. Conditional on self-reporting, Agarwal et al. (2013a) find significant evidence of both a timing bias and a delisting bias. Their results also indicate that self-reporting and nonreporting funds do not differ significantly in return performance.

⁸ We have three more strategies in the TASS database: (i) options strategy, (ii) other, and (iii) undefined. However, there are either very few observations for these strategies, or their investment methods or the maintained positions are not clear. Hence, we prefer to focus on the main 11 investment styles in the TASS database.

⁹ One-month T-bill returns are obtained from Kenneth French's online data library (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html). Ten-year Treasury returns are obtained from the Center for Research in Security Prices.

empirical return distributions are skewed, peaked around the mean, and leptokurtic with significant excess kurtosis. To formally test whether the hedge fund returns are normally distributed, we compute the Jarque–Bera (JB) statistics that strongly reject the null hypothesis of normality for all hedge fund strategies.¹⁰ The significance of high-order moments and JB statistics indicate that we need to take into account the differences in the entire return distributions of hedge fund and S&P 500 indices to determine which hedge fund strategies outperform the equity and bond markets.

Next, we consider the FSD criterion. For the hedge fund portfolio to dominate the S&P 500 index, the cumulative distribution of the hedge fund portfolio should always plot below the cumulative distribution of the S&P 500 index. One very practical way of observing whether this dominance criterion is violated is to look at the tails of the return distributions. If the minimum monthly return of the hedge fund portfolio is lower than that of the S&P 500 index, then the hedge fund index fails to dominate the S&P 500 index. For example, the emerging markets strategy has a staggering minimum monthly return of -22% compared with -17% for the S&P 500 index. Although the cumulative distribution of the emerging markets hedge fund index plots mostly below the cumulative distribution of the S&P 500 index, it plots above the cumulative distribution of the S&P 500 index for low return intervals. Similarly, the maximum returns of the convertible arbitrage, equity market neutral, event driven, fixed income arbitrage, fund of funds, global macro, long/short equity hedge, managed futures, and multistrategy hedge fund strategies are smaller than the maximum return (10.77%) of the S&P 500 index. Although short bias beats the S&P 500 index at the maximum and minimum values, it has a cumulative return distribution that plots above the cumulative distribution of the S&P 500 index for some other return intervals. Hence, none of the hedge fund indices performs better than the S&P 500 index in the sense of FSD.

Overall, these results indicate that although some of the hedge fund strategies outperform the S&P 500 index based on the mean-variance criterion, the distributional comparisons accounting for the higher-order moments do not point out the superior performance of hedge fund portfolios over the U.S. equity market index.

We now investigate whether the hedge fund strategies perform better than the short-term and long-

term fixed income securities. As reported in Table 2, the average returns on the one-month and 10-year Treasury securities are 0.26% and 0.55% per month. The corresponding standard deviations are 0.17% and 2.10% per month, respectively. Because both the average returns and standard deviations of the hedge fund indices are greater than the average returns and standard deviations of the one-month Treasury bill, there is no MV dominance of the hedge fund market over the short-term Treasury market. Interestingly, five hedge fund indices have mean-variance dominance over the long-term Treasury market: multi-strategy, event driven, global macro, equity market neutral, and fixed income arbitrage strategies have higher average returns and lower standard deviations compared with the 10-year Treasury returns. However, these five hedge fund indices do not dominate the 10-year Treasury in the sense of FSD because their maximum returns are lower than the maximum return of the 10-year Treasury bond (8.54%). Besides, their cumulative return distributions plot above the cumulative distribution of the 10-year Treasury bond for some other return intervals. Overall, we find that none of the hedge fund portfolios outperforms the one-month T-bill and 10-year T-bonds in the sense of FSD either.

3.3. Almost Dominance

The fact that the classical decision rules (MV and FSD) cannot identify the relative performance of hedge fund indices may have two reasons: (i) the practice of investing in hedge fund portfolios is not supported by the expected utility paradigm, or (ii) the limitations of these decision rules cause the failure of showing a dominance of some of the hedge fund strategies with more profitable distributional characteristics, even though almost all investors would choose them over the S&P 500 index and Treasury securities with relatively less favorable distributional characteristics.¹¹ Thus, we turn our attention to the ASD rules.

Hedge funds generally have an average lock-up period of one year in which investors are not allowed to redeem or sell shares. The lock-up period helps hedge fund portfolio managers avoid liquidity risk while capital is put to work in long investment horizons (see Aragon 2007). Furthermore, for longer investment periods, the cumulative distribution of the high-return portfolios may shift downward at a faster rate than the cumulative distribution of the low-return portfolios due to time diversification. Hence, even though there is no dominance at short investment horizons (such as one month), preference of

¹⁰ Another notable point in Table 2 is that the kurtosis is the highest for strategies such as convertible arbitrage, and fixed income arbitrage and is the lowest for managed futures. This may be due to the liquidity of the underlying securities hedge funds invest.

¹¹ The more profitable (or favorable) distributional characteristics imply relatively higher mean, higher skewness, lower standard deviation, and lower kurtosis.

some of the hedge fund portfolios at longer horizons may be justified within the expected utility paradigm. Therefore, we investigate the dominance of hedge fund indices for longer horizons. We compute long-term returns for the hedge fund and S&P 500 indices as well as the one-month and 10-year Treasury securities by cumulating monthly returns.¹²

We first examine the cumulative distributions of each hedge fund strategy and the S&P 500 index for investment horizons of one to five years. Specifically, we first determine the minimum and maximum return values of the two time series (say long/short equity hedge fund and S&P 500 indices). Then, we create a return bin, which consists of return values starting from the minimum return of the long/short equity hedge and S&P 500 indices going through the maximum return of the corresponding portfolios with 0.1% increments. Therefore, the size of this return bin depends on the minimum and the maximum values of the long/short equity hedge and S&P 500 indices and the precision parameter (i.e., size of the increments, which we hold fixed at 0.1%). After we create the return bin, we form the cumulative distributions by simply computing, for each portfolio, the percentage of times that a portfolio incurs a lower return than the corresponding return number in the bin. Thus, the number of data points in the cumulative distributions is equal to the size of the return bin. We should note that several other precision parameters generate very similar results.

For each comparison, we define A_1 as the estimated area between the cumulative distributions when the cumulative distribution of the long/short equity hedge fund portfolio with a higher mean plots above the cumulative distribution of the S&P 500 index with a lower mean (i.e., V in Figure 1). Similarly, we define A_2 as the estimated area between the cumulative distributions when the cumulative distribution of the S&P 500 index plots above the cumulative distribution of the long/short equity hedge fund portfolio (i.e., K in Figure 1). Then ε_1 is computed as the violation area (A_1) divided by the total area between the cumulative distributions of the long/short equity hedge fund and S&P 500 indices, i.e., $\varepsilon_1 = A_1 / (A_1 + A_2)$.

Panel A of Table 3 reports ε_1 values for each hedge fund strategy compared with the S&P 500 index for one- to five-year investment horizons. Panel A shows that for one-year horizon, ε_1 equals 0.16% for the

long/short equity hedge fund index, implying that the cumulative distribution of the long/short equity hedge fund portfolio is above the cumulative distribution of the S&P 500 index less than 1% of the time. As pointed out earlier, to obtain AFSD, the violation area must be lower than 5.9%. Because the empirical value of $\varepsilon_1 = 0.16\%$ is lower than the critical value of epsilon, $\varepsilon_1^* = 5.9\%$, the long/short equity hedge fund index dominates the S&P 500 index by AFSD at one-year horizon. The only other hedge fund strategy that dominates the U.S. equity market is the emerging markets hedge fund index with $\varepsilon_1 = 2.13\%$. Their superior performance continues for longer horizons as well. For two-, three-, four-, and five-year investment horizons, the long/short equity hedge and emerging markets hedge fund strategies beat the U.S. stock market. However, the remaining nine hedge fund strategies do not generate superior performance over the S&P 500 index in the sense of AFSD.¹³

Another interesting observation in Table 3 is that although ε_1 values are greater than the critical value of 5.9% for managed futures and multistrategy, they do fairly well compared with the nondirectional or semidirectional strategies such as equity market neutral, convertible arbitrage, and fixed income arbitrage. Overall, the results indicate that moving from nondirectional/semidirectional to strongly directional strategies, the probability of those specific hedge fund strategies dominating U.S. equity markets increases.

Panel B of Table 3 presents ε_1 values for each hedge fund strategy compared with the one-month Treasury bills for one- to five-year investment horizons. Panel B shows that for one-year horizon, ε_1 equals 1.31% for the global macro hedge fund index, 2.22% for the managed futures, 3.44% for the multistrategy, and 5.80% for the long/short equity hedge, all of which are lower than the critical value $\varepsilon_1^* = 5.9\%$. These results indicate that at one-year horizon, the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies outperform the one-month T-bills based on the AFSD. For two-year investment horizon, in addition to the global macro, managed futures, multistrategy, and long/short equity, the fixed income arbitrage, equity market neutral, and event driven hedge fund indices also beat the short-term Treasury market. For three-, four-, and five-year investment horizons, all hedge fund strategies (with the only exception of short bias)

¹² As will be discussed in this section, hedge fund strategies become gradually stronger as we move from a one-year to a five-year investment horizon. This suggests that the time diversification effects are greater for hedge funds as compared with stocks and bonds. In fact, this effect is even more pronounced for illiquid style categories (e.g., emerging markets), where share restrictions are more likely to be used.

¹³ In the online appendix, we investigate the relative performance of the long/short equity hedge and emerging markets strategies against the remaining nine hedge fund strategies during up and down markets. Stock market upturns and downturns are defined based on the median S&P 500 index return, observations above (below) the median are denoted as *up market* (*down market*). The results indicate superior performance of the long/short equity hedge during rises of the market and superior performance of emerging markets during large falls of the market.

Table 3 Almost First-Order Stochastic Dominance

Hedge fund portfolio	Panel A: Hedge funds vs. S&P 500 Index					Panel B: Hedge funds vs. one-month T-bill					Panel C: Hedge funds vs. 10-year T-bond				
	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year
Convertible arbitrage	0.4269	0.4232	0.4202	0.4079	0.3393	0.1795	0.0950	0.0535	0.0376	0.0399	0.2659	0.2078	0.1951	0.2282	0.1835
Dedicated short bias	0.8119	0.8159	0.8139	0.8159	0.8665	0.5331	0.5798	0.6292	0.7826	0.9312	0.7770	0.8498	0.9276	0.9999	0.9999
Emerging markets	0.0213	0.0000	0.0042	0.0000	0.0000	0.2055	0.1019	0.0074	0.0101	0.0000	0.2430	0.1521	0.0634	0.0363	0.0068
Equity market neutral	0.4429	0.4128	0.3544	0.3196	0.2573	0.0664	0.0207	0.0000	0.0000	0.0000	0.0536	0.0830	0.0868	0.0790	0.0611
Event driven	0.3131	0.3084	0.2947	0.2713	0.2139	0.1033	0.0525	0.0083	0.0000	0.0000	0.1074	0.1005	0.0699	0.0441	0.0066
Fixed income arbitrage	0.4524	0.4259	0.3909	0.3871	0.3437	0.0650	0.0262	0.0037	0.0000	0.0000	0.0365	0.0716	0.0733	0.1014	0.0797
Fund of funds	0.5601	0.5444	0.4994	0.4712	0.4112	0.1777	0.1134	0.0513	0.0439	0.0118	0.4992	0.3749	0.3642	0.3233	0.2304
Global macro	0.4028	0.3727	0.3238	0.3135	0.2713	0.0131	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Long/short equity hedge	0.0016	0.0000	0.0026	0.0005	0.0000	0.0580	0.0195	0.0012	0.0001	0.0000	0.0517	0.0462	0.0186	0.0049	0.0009
Managed futures	0.3721	0.3747	0.3277	0.3120	0.2742	0.0222	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Multistrategy	0.2950	0.2543	0.2160	0.1900	0.1261	0.0344	0.0001	0.0000	0.0000	0.0000	0.0040	0.0013	0.0000	0.0000	0.0000

Notes. This table presents the empirical estimates of ε_1 for one- to five-year investment horizons. For each comparison, we define A_1 as the area between the cumulative distributions when the cumulative distribution of a hedge fund portfolio plots above the cumulative distribution of the S&P 500 index or Treasury security (V in Figure 1). Similarly, we define A_2 as the area between the cumulative distributions when the cumulative distribution of the S&P 500 index (or Treasury security) plots above the cumulative distribution of the hedge fund portfolio (K in Figure 1). The measure of ε_1 for AFSD is defined as $A_1/(A_1 + A_2)$ and is calculated using the actual data for the sample period January 1994–December 2011. The critical value for ε_1 is $\varepsilon_1^* = 5.9\%$ for AFSD. The results are presented based on the comparison of actual empirical return distributions. Panels A, B, and C present results from comparing the hedge fund portfolios with the S&P 500 index, one-month Treasury bill, and 10-year Treasury bond, respectively.

perform better than the one-month Treasury bill in the sense of AFSD.

Panel C of Table 3 repeats our analysis for 10-year Treasury bonds. Panel C shows that for all investment horizons, the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies outperform the 10-year Treasury bond based on the AFSD. For the one-year investment horizon, in addition to the aforementioned four strategies, the equity market neutral and fixed income arbitrage also beat the long-term Treasury market. For four- and five-year investment horizons, most hedge fund strategies (with the exception of short bias, fund of funds, fixed income, convertible arbitrage, and equity market neutral) perform better than the 10-year Treasury security in the sense of AFSD.

Next, we present results from the ASSD analysis. For each distributional comparison, we define A_3 as the estimated area between the cumulative distributions of the long/short equity hedge fund and S&P 500 indices, when the cumulative distribution of the long/short equity hedge fund portfolio plots above such that the aggregated area between the two distributions becomes positive (i.e., Q in Figure 2). The area A_4 is the estimated total absolute area enclosed between the two distributions. Then ε_2 is computed as the ratio of the violation area (A_3) to the total absolute area, i.e., ε_2 is computed using Equation (6), $\varepsilon_2 = A_3/A_4$.

As discussed earlier, ε_2 values help us determine the ASSD of hedge fund portfolios. Panel A of Table 4 shows that even though most of the hedge fund strategies achieve ASSD over the S&P 500 index for short-term investment horizons of one and two years, there

is no ASSD dominance of the fund of funds and short bias. Specifically, ε_2 values are estimated to be higher than the critical value of $\varepsilon_2^* = 3.2\%$ for the fund of funds and short bias for one- and two-year horizons. For three-, four-, and five-year horizons, all hedge fund strategies (with the only exception of short bias) outperform the U.S. equity market in the sense of ASSD. Panel B of Table 4 shows that at the one-year horizon, only the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies outperform the one-month T-bills based on the ASSD. The remaining seven hedge fund strategies do not outperform the short-term Treasury security at one-year horizon. For long investment horizons of four and five years, all hedge fund strategies have ASSD dominance over the one-month Treasury bills, with the exception of short bias. Panel C of Table 4 reports that for all investment horizons, the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies outperform the 10-year Treasury bond based on the ASSD. However, the remaining seven hedge fund indices do not have ASSD dominance over the long-term Treasury market for one-, two-, and three-year investment horizons. For long-term investment horizons of four and five years, in addition to the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies, the emerging markets and event driven strategies have success in terms of outperforming the 10-year Treasury in the sense of ASSD.

3.4. Confidence Bands for the AFSD Statistics

We have so far provided evidence from the realized empirical return distributions. In the online appendix (available at <http://faculty.msb.edu/tgb27/>),

Table 4 Almost Second-Order Stochastic Dominance

Hedge fund portfolio	Panel A: Hedge funds vs. S&P 500 Index					Panel B: Hedge funds vs. one-month T-bill					Panel C: Hedge funds vs. 10-year T-bond				
	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year
Convertible arbitrage	0.0000	0.0000	0.0000	0.0000	0.0000	0.1795	0.0950	0.0535	0.0316	0.0312	0.2659	0.2078	0.1951	0.2282	0.1835
Dedicated short bias	0.6262	0.6335	0.6288	0.6322	0.7331	0.5331	0.5798	0.6292	0.7826	0.9312	0.7770	0.8498	0.9276	0.9999	0.9999
Emerging markets	0.0080	0.0000	0.0000	0.0000	0.0000	0.2055	0.1019	0.0074	0.0101	0.0000	0.2430	0.1521	0.0634	0.0313	0.0068
Equity market neutral	0.0000	0.0000	0.0000	0.0000	0.0000	0.0664	0.0207	0.0000	0.0000	0.0000	0.0387	0.0830	0.0860	0.0790	0.0611
Event driven	0.0000	0.0000	0.0000	0.0000	0.0000	0.1033	0.0525	0.0083	0.0000	0.0000	0.1074	0.1005	0.0699	0.0319	0.0066
Fixed income arbitrage	0.0000	0.0000	0.0000	0.0000	0.0000	0.0650	0.0262	0.0037	0.0000	0.0000	0.0365	0.0716	0.0733	0.1014	0.0797
Fund of funds	0.1215	0.0893	0.0000	0.0000	0.0000	0.1777	0.1134	0.0513	0.0319	0.0118	0.4992	0.3749	0.3642	0.3233	0.2304
Global macro	0.0000	0.0000	0.0000	0.0000	0.0000	0.0131	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Long/short equity hedge	0.0000	0.0000	0.0000	0.0000	0.0000	0.0317	0.0195	0.0012	0.0001	0.0000	0.0314	0.0262	0.0186	0.0049	0.0009
Managed futures	0.0000	0.0000	0.0000	0.0000	0.0000	0.0222	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Multistrategy	0.0000	0.0000	0.0000	0.0000	0.0000	0.0314	0.0001	0.0000	0.0000	0.0000	0.0032	0.0000	0.0000	0.0000	0.0000

Notes. This table presents the empirical estimates of ε_2 for one- to five-year investment horizons. For each comparison, we define A_3 as the estimated area between the cumulative distributions of a hedge fund and S&P 500 indices (or Treasury security), when the cumulative distribution of the hedge fund portfolio plots above such that the aggregated area between the two distributions becomes positive (i.e., Q in Figure 2). We define A_4 as the estimated total absolute area enclosed between the two distributions. Then ε_2 is computed as the ratio of the violation area (A_3) to the total absolute area, i.e., $\varepsilon_2 = A_3/A_4$ and is calculated using the actual data for the sample period January 1994–December 2011. The critical value for ε_2 is $\varepsilon_2^* = 3.2\%$ for ASSD. The results are presented based on the comparison of actual empirical return distributions. Panels A, B, and C present results from comparing the hedge fund portfolios with the S&P 500 index, one-month Treasury bill, and 10-year Treasury bond, respectively.

we analyze in depth cumulative distributions of hedge fund indices based on simulated return series. In this section, we show how to use this simulation evidence to estimate confidence bands for the AFSD statistics. We focus on the two main series: long/short equity hedge and emerging markets. The actual data are based on 216 monthly observations. For k -month horizon ($k = 12$ to 60 months), we generate the simulated data by randomly picking k monthly observations 2,000 times and compounding these k observations to generate one-year to five-year returns. These simulated series are used to obtain simulated ε_1 values. This process is repeated 1,000 times, and, hence, for each horizon, 1,000 ε_1 values are generated. Because we rely on the simulations of the simulated ε_1 values, considerable processing power is required. For example, to compare two return series at a 60-month horizon and to obtain 1,000 ε_1 values, we use 120 million return observations for each series.

For each horizon, the distribution of 1,000 ε_1 values is generated, and the mean and standard deviation of the estimated ε_1 values are computed. In addition, 1% confidence intervals for the AFSD statistics are obtained by using the bootstrap estimate of standard errors of the 1,000 ε_1 values, i.e., the observed standard deviation of repeated bootstrap replications without any scaling. Panel A of Table IV of the online appendix presents these statistics to compare the performance of the long/short equity hedge with the S&P 500 index, and panel B of Table IV reports these statistics to compare the emerging markets strategy with the S&P 500 index. Table IV shows that ε_1 estimates are extremely precise with highly restricted confidence bands. For example, for three-year investment horizon, the long/short equity hedge dominates the S&P 500 index with a mean simulated epsilon value

of 2.27%, and the 1% confidence interval of this ε_1 value is [2.15%, 2.38%]. Similarly, for the three-year investment horizon, the emerging markets strategy dominates the S&P 500 index with a mean simulated epsilon value of 0.05%, and the corresponding 1% confidence interval is [0.04%, 0.06%]. We conclude that very accurate estimates of the ASD statistics with highly narrow confidence intervals are obtained for all investment horizons.

4. Traditional and Manipulation-Proof Performance Measures

Goetzmann et al. (2007) point out the vulnerability of the traditional performance measures to a number of simple dynamic manipulation strategies and develop a manipulation-proof performance measure:

MPPM

$$= \frac{1}{(1-\lambda)\Delta t} \ln \left(\frac{1}{T} \sum_{t=1}^T [(1+R_{p,t})/(1+r_{f,t})]^{1-\lambda} \right), \quad (7)$$

where $R_{p,t}$ and $r_{f,t}$ are the portfolio return and the risk-free interest rate at time t . Goetzmann et al. (2007) transform the measure so that Equation (7) can be interpreted as the annualized continuously compounded excess return certainty equivalent of the portfolio. They relate the MPPM with some benchmark portfolio that is chosen to be the market index. They show that if the benchmark market portfolio has a lognormal return, $1+r_{m,t}$, then the parameter λ should be selected so that

$$\lambda = \frac{\ln[E(1+r_m)] - \ln[E(1+r_f)]}{\text{Var}[1+r_m]}. \quad (8)$$

In our empirical analyses, we allow λ to change for different data frequencies, and we keep $\Delta t = 1$.

In Equation (8), the market return, r_m , is proxied by the S&P 500 index, and the risk-free rate, r_f , is proxied by the one-month Treasury bill.

The MPPM of Goetzmann et al. (2007) is a sophisticated parametric approach and, similar to ASD, is also utility-based. Hence, we test whether the utility-based parametric measure of MPPM and the utility-based nonparametric measure of ASD generate similar rankings for the relative performance of hedge fund strategies.

Panel A of Table 5 presents the MPPMs for the hedge fund strategies and the S&P 500 index for the sample period January 1994–December 2011. A notable point in Table 5 is that for all investment horizons, the long/short equity hedge and emerging markets strategies produce the highest MPPMs compared with the other hedge fund strategies and the S&P 500 index. Specifically, for the short-term horizon of one year, the MPPMs for the long/short equity hedge and emerging markets hedge fund portfolios are, respectively, 0.0911 and 0.0789, whereas the MPPM for the S&P 500 index is only 0.0244. For the long-term horizon of five years, the MPPMs for the long/short equity hedge and emerging markets styles are, respectively, 0.4688 and 0.4906, whereas the MPPM for the S&P 500 index is only 0.0622. Consistent with our earlier findings from the ASD approach, the short bias, equity market neutral, fixed income arbitrage, convertible arbitrage, and fund of funds are consistently among the worst performers for all investment horizons. Another notable point in Table 5, panel A, is that for all investment horizons, the multistrategy is the closest follower of the long/short equity hedge and emerging markets indices. The MPPMs for the long/short equity hedge are in the range of 0.0911 to 0.4688, whereas the MPPMs for the multistrategy range from 0.0770 to 0.4065. Overall, the results indicate that the directional hedge funds with dynamic trading strategies produce the highest MPPMs, whereas the least directional hedge fund strategies are among the worst performers.¹⁴

¹⁴ Bali et al. (2012) show that hedge fund managers deliver significant alpha by increasing or decreasing their portfolios' aggregate exposure to risk factors by timely predicting the changes in financial and economic conditions. They find that hedge funds following directional, dynamic trading strategies correctly adjust their aggregate exposure to changes in risk factors, whereas this type of skill is low for funds following nondirectional strategies, such as equity market neutral, fixed income arbitrage, and convertible arbitrage funds. Consistent with Bali et al. (2012), we find that the directional strategies outperform the nondirectional investment styles. Agarwal et al. (2013b) uncover hedge fund skill from the portfolio holdings they hide. After examining the confidential holdings of hedge funds, they find that funds managing large risky portfolios with nonconventional strategies seek confidentiality more frequently, and their portfolio holdings exhibit superior performance up to one year.

Table 5 Manipulation-Proof Performance and Traditional Measures

Assets	Panel A: MPPMs					Panel B: Sharpe ratios					Panel C: Treynor ratios					Panel D: Monthly alphas		
	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year	7-factor alpha (%)	t-stat.	
Convertible arbitrage	0.047	0.099	0.135	0.170	0.219	0.475	0.722	0.904	0.956	1.038	0.143	0.395	0.701	0.940	1.159	0.34	(1.99)	
Dedicated short bias	-0.016	-0.029	-0.036	-0.051	-0.081	-0.045	-0.101	-0.153	-0.308	-0.559	0.013	0.056	0.142	0.356	0.706	0.48	(2.12)	
Emerging markets	0.079	0.180	0.274	0.367	0.491	0.515	0.823	1.008	1.129	1.267	0.221	0.966	4.429	-9.902	-6.308	0.55	(1.43)	
Equity market neutral	0.054	0.109	0.160	0.214	0.268	0.921	1.043	1.071	1.138	1.213	0.284	0.528	0.697	0.928	1.144	0.47	(7.00)	
Event driven	0.068	0.138	0.199	0.266	0.343	0.808	1.150	1.367	1.570	1.891	0.196	0.495	0.851	1.314	1.925	0.45	(4.12)	
Fixed income arbitrage	0.052	0.106	0.151	0.192	0.240	0.812	1.011	1.130	1.307	1.534	0.346	0.723	0.998	1.379	1.842	0.47	(4.54)	
Fund of funds	0.031	0.069	0.104	0.142	0.189	0.447	0.672	0.820	0.975	1.152	0.138	0.359	0.564	0.886	1.255	0.25	(2.32)	
Global macro	0.062	0.129	0.191	0.247	0.304	1.057	1.501	1.769	2.139	2.526	0.328	0.727	0.992	1.427	1.940	0.49	(5.11)	
Long/short equity hedge	0.091	0.188	0.278	0.369	0.469	0.789	1.040	1.136	1.216	1.253	0.204	0.449	0.645	0.869	1.013	0.61	(5.58)	
Managed futures	0.064	0.128	0.196	0.262	0.321	0.975	1.616	2.177	3.099	4.317	-3.946	4.485	2.885	3.789	4.731	0.87	(5.25)	
Multistrategy	0.077	0.161	0.239	0.320	0.407	1.160	1.590	1.790	1.911	2.017	0.342	0.796	1.187	1.670	2.054	0.60	(6.90)	
S&P 500	0.024	0.048	0.056	0.061	0.062	0.249	0.302	0.280	0.274	0.271	0.048	0.099	0.129	0.163	0.185	—	—	

Notes: Panels A, B, and C present the MPPMs, the Sharpe ratios, and the Treynor ratios for the hedge fund strategies and the S&P 500 index for the sample period January 1994–December 2011. Panel D reports the seven-factor monthly alphas for the hedge fund portfolios and their Newey–West t -statistics in parentheses.

As discussed earlier, the ASD approach of Leshno and Levy (2002) does not require a parametric specification of investors' preferences and does not make any assumptions about asset returns. Instead, ASD imposes general conditions regarding preferences and considers the entire return distribution with high-order moments. Similar to the ASD approach, the MPPM of Goetzmann et al. (2007) is a utility-based performance measure. Overall, the results in panel A of Table 5 provide strong evidence that the utility-based parametric measure of MPPM and the utility-based nonparametric measure of ASD generate very similar rankings for the relative performance of hedge fund strategies.

Now we investigate whether the traditional measures produce different rankings for the relative strength of hedge fund strategies. We use three standard measures: (i) the Sharpe ratio, (ii) the Treynor ratio, and (iii) the seven-factor alpha. The Sharpe ratio, defined as the excess return (expected return in excess of the risk-free rate) per unit of standard deviation, is used to characterize how well the return of an asset compensates the investor for the risk taken. The Treynor ratio is defined as the excess return per unit of market beta. Whereas the Treynor ratio works only with systemic risk of a portfolio, the Sharpe ratio observes both systemic and idiosyncratic risks.

The Sharpe and Treynor ratios have been challenged with regard to their appropriateness as fund performance measures during evaluation periods of declining markets. As long as the returns are normally distributed, these traditional measures can be useful for ranking individual assets or portfolios. However, not all asset or portfolio returns are normally distributed. Abnormalities like skewness, kurtosis, fatter tails, and higher peaks of the distribution can be problematic for these ratios, as standard deviation does not have the same effectiveness when these nonnormalities exist.

Panels B and C of Table 5 report the Sharpe and Treynor ratios for the hedge fund strategies and the S&P 500 index for the period 1994–2011. A first notable point in panels B and C is that the portfolio rankings based on the Sharpe and Treynor ratios are sensitive to the investment horizon. Another notable point is that depending on the investment horizon, the long/short equity hedge and emerging markets strategies are ranked seventh or eighth based on the Sharpe ratios of 11 hedge fund strategies, although the long/short equity hedge and emerging markets are consistently the best performers based on the ASD measure and MPPM. Likewise, based on the comparisons of Treynor ratios in panel C, the long/short equity hedge is ranked in the range of 6th and 9th, and the emerging markets generally fluctuates between 5th and 11th depending the investment horizon. Similar inconsistencies exist in panels B and C for the other hedge fund portfolios as well.

The most important feature of panels B and C is that the directional funds with dynamic trading strategies (e.g., the long/short equity hedge and emerging markets) are ranked among the worst performers, whereas the least directional funds with inactive trading strategies are ranked among the best performers according to these traditional measures.

In addition to the Sharpe and Treynor ratios, we estimate the monthly alphas of the hedge fund portfolios based on the seven-factor model of Fung and Hsieh (2004). Specifically, we use the S&P 500 index as a proxy for the equity market factor, the SMB (small-minus-big) factor of Fama and French (1993) as a proxy for the size factor, and the five trend-following factors of Fung and Hsieh (2001) to proxy for the nonlinear asymmetric payoff patterns of the hedge fund strategies.¹⁵ The excess monthly returns of hedge fund indices are regressed on a constant and the seven factors. The magnitudes of seven-factor alphas reported in panel D of Table 5 indicate that the top four performers are the managed futures, the long/short equity hedge, multistrategy, and emerging markets strategies, with respective alphas of 0.87%, 0.61%, 0.60%, and 0.55% per month. A first notable point in panel D is that the long/short equity hedge and emerging markets strategies are not the best performers based on the seven-factor alphas. In addition to having managed futures as the best performer, the multistrategy performs almost as well as the long/short equity hedge and even better than the emerging market index. Another notable point is that although the emerging markets style is ranked fourth among the 11 hedge fund strategies, the seven-factor alpha (0.55% per month) for the emerging markets is statistically insignificant, with a very low Newey–West *t*-statistic of 1.43. Interestingly, the short bias, equity market neutral, and fixed income arbitrage strategies—among the worst performers according to the ASD measure and MPPM—are ranked right after the long/short equity hedge and emerging markets strategies with statistically significant alphas. Panel D also shows that the managed futures, multistrategy, and global macro strategies that consistently outperform the U.S. Treasury market and generate similar ASD measures and MPPMs in our earlier tables are now differentiated by the seven-factor alphas.

Although the performance rankings obtained from the seven-factor alphas turn out to be closer to

¹⁵ The five trend-following factors of Fung and Hsieh (2001) are (i) the return of bond lookback straddle, (ii) the return of currency lookback straddle, (iii) the return of commodity lookback straddle, (iv) the return of short-term interest rate lookback straddle, and (v) the return of stock index lookback straddle. These trend-following factors are provided by David Hsieh at <http://faculty.fuqua.duke.edu/~dah7/HFRFData.htm> (accessed August 2012).

the ASD measures and MPPMs (compared with the Sharpe and Treynor ratios), we still observe some inconsistencies in the alpha comparisons.¹⁶ The results in Table 5 indicate that because the return distributions exhibit significant departures from normality, the Sharpe and Treynor ratios as well as the alphas do not provide an accurate characterization of the relative strength of hedge fund portfolios.

5. Downside Risk-Adjusted Performance Measures

The mean-variance analysis developed by Markowitz (1952) critically relies on two assumptions: either the investors have a quadratic utility or the asset returns are jointly normally distributed. Both assumptions are not required, just one or the other: (i) If an investor has quadratic preferences, she cares only about the mean and variance of returns, but she will not care about extreme losses, and (ii) mean-variance optimization can be justified if the asset returns are jointly normally distributed because the mean and variance will completely describe the distribution. However, as discussed earlier, the empirical distribution of hedge fund returns is skewed and has fat tails, implying that extreme events occur much more frequently than predicted by the Normal distribution. Therefore, the traditional measures of market risk are not appropriate to estimate the maximum likely loss that a fund can expect to lose under highly volatile periods.

There is a long literature about safety-first investors who minimize the chance of disaster (or the probability of failure). The portfolio choice of a safety-first investor is to maximize expected return subject to a downside risk constraint. Roy's (1952), Baumol's (1963), Levy and Sarnat's (1972), and Arzac and Bawa's (1977) safety-first investor uses a downside risk measure that is a function of value at risk (VaR). Following these studies, we use a modified Sharpe ratio that replaces standard deviation with the 1% VaR:

$$Mean - VaR = \frac{E(R_p) - r_f}{VaR}, \quad (9)$$

where $E(R_p) - r_f$ is the expected excess return on a hedge fund portfolio, and the 1% VaR is computed

using the left tail of the empirical return distribution (one percentile of the actual return distribution).

Panel A of Table 6 shows that when VaR is used as a measure of risk in the modified Sharpe ratio, the long/short equity hedge and emerging markets strategies are ranked between fourth and eighth for one- to three-year investment horizon. Even though, for four-year investment horizon, the long/short equity hedge is the best performer, the emerging markets strategy is ranked eighth among 11 hedge fund strategies. For five-year investment horizon, the long/short equity hedge and emerging markets strategies are ranked third and fifth, respectively. Another notable point in Table 6 is that for all investment horizons, almost all hedge fund strategies (with the exception of short bias) outperform the S&P 500 index. Hence, replacing standard deviation with VaR does not generate a robust, consistent ranking among alternative investment styles.

In addition to the mean-VaR ratio, we use an alternative risk-adjusted measure of performance based on the maximum drawdown. The maximum cumulative loss from a market peak to the following trough, often called the maximum drawdown (MDD), is a measure of how sustained one's losses can be. Large drawdowns usually lead to fund redemptions, and so the MDD is the risk measure of choice for many fund managers—a reasonably low MDD is critical to the success of any fund. The Calmar ratio is a downside risk-adjusted performance measure based on MDD:

$$Calmar = \frac{E(R_p) - r_f}{MDD}, \quad (10)$$

where MDD is computed as the largest drawdown for one- to five-year investment horizons during the sample period 1994–2011.

The results from the Calmar ratio are similar to those obtained from the ASD measure and MPPM. Panel B of Table 6 shows that when MDD is used as a measure of downside risk, the long/short equity hedge and emerging markets strategies are consistently the best performers among 11 hedge fund portfolios. According to the Calmar ratio in Equation (10), the worst performers are the short bias, fixed income arbitrage, equity market neutral, and fund of funds strategies. The maximum drawdown indicates the greatest loss an investor could suffer if an investment is bought at its highest price and sold at the lowest. Hence, the Calmar ratio is considered a good indicator of the emotional pain an investor could feel if the fund suddenly swings downward.

Finally, we use the Sortino ratio based on the lower partial second moment (LPM_2). The Sortino ratio (named after Frank A. Sortino), measures the downside risk-adjusted return of an investment. It is a modification of the Sharpe ratio, but penalizes only

¹⁶ Agarwal and Naik (2000) point out that hedge funds' extensive use of leverage can sometimes obscure alpha measures. They show that alpha is inappropriate for hedge funds because it is not invariant to leverage. Instead, they argue for the appraisal ratio defined as the ratio of alphas to the standard error of the factor model regressions (see Brown et al. 1992). The literature relies on the Treynor ratio and the Treynor appraisal ratio to negate the leverage effect on alpha. Although the Treynor measure has some undesirable statistical properties (in repeated samples as the ratio of asymptotically Normal variates with nonzero means it is distributed as Normal-Cauchy and has no finite moments), it is at least invariant to leverage.

Table 6 Downside Risk-Adjusted Performance Measures

Assets	Panel A: Mean-VaR ratio					Panel B: Calmar ratio					Panel C: Sortino ratio				
	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year	1-year	2-year	3-year	4-year	5-year
Convertible arbitrage	0.168	0.343	0.520	0.646	0.897	0.348	0.184	0.138	0.114	0.096	0.903	1.685	2.533	3.336	4.231
Dedicated short bias	−0.024	−0.055	−0.072	−0.126	−0.178	−0.109	−0.073	−0.056	−0.047	−0.044	−0.068	−0.150	−0.211	−0.323	−0.479
Emerging markets	0.263	0.685	2.450	2.321	5.444	0.465	0.239	0.170	0.136	0.111	0.977	2.423	8.626	9.916	31.32
Equity market neutral	0.591	1.418	2.724	8.968	24.00	0.337	0.178	0.128	0.103	0.088	2.676	6.420	14.22	52.80	1,974.6
Event driven	0.352	0.714	1.559	3.891	21.95	0.376	0.198	0.144	0.116	0.097	1.732	3.314	7.836	23.60	2,536.5
Fixed income arbitrage	0.441	0.846	1.431	2.433	8.709	0.333	0.174	0.128	0.105	0.090	2.270	4.245	7.078	12.30	50.60
Fund of funds	0.205	0.497	0.882	1.461	3.821	0.258	0.134	0.096	0.077	0.065	0.883	1.820	3.027	5.558	17.32
Global macro	1.241	17.35	4.170	3.223	2.558	0.349	0.181	0.129	0.106	0.092	5.916	72.76	1,165.3	1,617.7	2,153.9
Long/short equity hedge	0.562	1.542	3.161	12.76	10.78	0.420	0.221	0.159	0.128	0.109	2.875	7.249	18.26	269.70	3,955.3
Managed futures	1.316	10.90	8.144	2.568	1.979	0.358	0.190	0.133	0.106	0.092	5.416	36.55	1,189.5	1,710.7	2,267.3
Multistrategy	0.817	3.545	13.92	3.731	3.025	0.380	0.196	0.140	0.112	0.095	4.542	19.45	743.80	2,209.4	3,105.4
S&P 500	0.116	0.201	0.247	0.326	0.412	0.343	0.191	0.146	0.126	0.109	0.392	0.561	0.636	0.775	0.928

Notes: Panels A, B, and C present the downside risk-adjusted performance measures (mean-VaR, Calmar, and Sortino ratios) that replace standard deviation in the Sharpe ratio with the 1% VaR, MDD, and LPM_2 , respectively. The results are reported for the hedge fund strategies and the S&P 500 index for the sample period January 1994–December 2011.

those returns falling below a user-specified target, or required rate of return, whereas the Sharpe ratio penalizes both upside and downside volatility equally. The Sortino ratio is calculated as

$$\text{Sortino} = \frac{E(R_p) - r_f}{LPM_2}, \quad (11)$$

$$LPM_2 = \int_{-\infty}^h (R - h)^2 f(R) dR, \quad (12)$$

where h is the target level of returns, and $f(R)$ represents the probability density function of returns. The measure LPM_2 is built on the notion of semivariance and the lower partial moment of Markowitz (1959). The semivariance is the expected value of the squared negative deviations of the possible outcomes from the mean, whereas the more general LPM_2 uses a chosen point of reference (h). The main heuristic motivation for the use of the LPM in place of variance as a measure of risk is that the LPM measures losses (relative to some reference point), whereas variance depends on gains as well as losses. In our empirical analysis, the target return h (or a minimum acceptable return) is assumed to be the risk-free rate. Lower Sortino ratios signify investments with a greater risk of large losses and should be avoided by risk-averse investors.

Similar to our findings from the mean-VaR ratio, replacing the standard deviation with LPM_2 does not generate a robust, consistent ranking among hedge fund strategies. Panel C of Table 6 shows that for investment horizons of one, two, and four years, the long/short equity hedge and emerging markets strategies are ranked between fourth and eighth. For the three-year investment horizon, the long/short equity hedge and emerging markets strategies are ranked fourth and sixth, respectively. However, for the five-year investment horizon, the long/short equity hedge is the best performer, whereas the emerging markets style is ranked eighth. According to the Sortino ratio in Equation (11), the worst performers are generally the short bias, convertible arbitrage, and fund of funds strategies. Similar to our findings from the mean-VaR ratio, all hedge fund strategies (with the exception of short bias) outperform the S&P 500 index for all investment horizons.

Overall, the results in Table 6 indicate that the downside risk-adjusted performance measures (mean-VaR, Calmar, and Sortino ratios) produce conflicting evidence for the relative performance of hedge fund strategies. In other words, the general conclusion is sensitive to the choice of a downside risk measure.

6. Time-Varying Conditional Distributions

We have so far investigated the realized empirical return distributions. However, in practice investors

or fund managers are not able to observe the future return distributions. Instead, they form expectations about possible outcomes on their portfolios using some state variables in their information set. Hence, in essence the relative performance of hedge funds, stocks, and bonds depends on the conditionally expected return distributions. In this section, we examine the relative strength of hedge fund strategies with reference to equity and bond markets using time-varying conditional distributions. There is substantial evidence that the distribution of financial asset returns is skewed and shows high peaks and fat tails. To account for skewness and kurtosis in the data, we use the skewed generalized error distribution (SGED) of Bali and Theodossiou (2008) that takes into account the nonnormality of returns and relatively infrequent events. The probability density function for the SGED is

$$f(r_t; \mu, \sigma, k, \lambda) = \frac{C}{\sigma} \exp\left(-\frac{1}{[1 + \text{sign}(r_t - \mu + \delta\sigma)\lambda]^k \theta^k \sigma^k} |r_t - \mu + \delta\sigma|^k\right), \quad (13)$$

where $C = k/(2\theta\Gamma(1/k))$; $\theta = \Gamma(1/k)^{0.5}\Gamma(3/k)^{-0.5}S(\lambda)^{-1}$; $\delta = 2\lambda AS(\lambda)^{-1}$; $S(\lambda) = \sqrt{1 + 3\lambda^2 - 4A^2\lambda^2}$; $A = \Gamma(2/k)\Gamma(1/k)^{-0.5}\Gamma(3/k)^{-0.5}$; μ and σ are the mean and standard deviation of returns r_t , respectively; λ is a skewness parameter; sign is the sign function; and $\Gamma(\cdot)$ is the gamma function. The scaling parameters k and λ obey the following constraints: $k > 0$ and $-1 < \lambda < 1$. The parameter k controls the height and tails of the density function, and the skewness parameter, λ , controls the rate of descent of the density around the mode of r , where $\text{mode}(r) = \mu - \delta\sigma$: $\lambda > 0$ implies that the density function is skewed to the right (positive skewness), whereas $\lambda < 0$ implies that the density function is skewed to the left (negative skewness).

To take into account skewness and tail thickness as well as the time-series variation in the moments of the return distribution, we use the conditional SGED density with time-varying conditional mean and volatility:

$$r_t = \mu_{t|t-1} + \varepsilon_t, \quad (14)$$

$$E(\varepsilon_t^2 | \Omega_{t-1}) = \sigma_{t|t-1}^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \sigma_{t-1}^2, \quad (15)$$

where r_t is the return that follows the conditional SGED density, $\mu_{t|t-1}$ is the conditional mean, and $\sigma_{t|t-1}^2$ is the conditional variance of returns based on the information set up to time $t - 1$, Ω_{t-1} . The conditional variance measure $\sigma_{t|t-1}^2$ in Equation (15) is the GARCH (generalized autoregressive conditional heteroscedasticity) model, originally proposed

by Bollerslev (1986), which defines the current conditional volatility as a function of the last period's unexpected news and the last period's volatility. The conditional mean $\mu_{t|t-1}$ in Equation (14) is parameterized as a function of the one-month lagged return, r_{t-1} , i.e., an AR(1) process:

$$\mu_{t|t-1} = \alpha_0 + \alpha_1 r_{t-1}. \quad (16)$$

The parameters of the conditional SGED density are estimated by maximizing the conditional log-likelihood function of r_t with respect to the parameters in $\mu_{t|t-1}$ and $\sigma_{t|t-1}^2$, and the skewness (λ) and tail-thickness (k) parameters:

$$\begin{aligned} \text{Log } L(\mu_{t|t-1}, \sigma_{t|t-1}, k, \lambda) \\ = n \ln C - n \ln \sigma_{t|t-1} \\ - \frac{1}{\theta^k \sigma_{t|t-1}^k} \sum_{t=1}^n \left(\frac{|r_t - \mu_{t|t-1} + \delta \sigma_{t|t-1}|^k}{[1 + \text{sign}(r_t - \mu_{t|t-1} + \delta \sigma_{t|t-1})\lambda]^k} \right), \end{aligned} \quad (17)$$

where C , θ , and δ are defined below Equation (13), sign is the sign of the residuals ($r_t - \mu_{t|t-1} + \delta \sigma_{t|t-1}$), and n is the sample size.

In this section, we compare the relative performance of the long/short equity hedge and emerging markets strategies with the U.S. equity and bond markets. We first estimate the parameters of the conditional SGED density for the long/short equity hedge and emerging markets styles and the S&P 500 index, one-month, and 10-year Treasury returns for the sample period January 1994–December 2011. Once we generate the conditional distributions, we investigate the relative strength of the hedge fund styles for investment horizons of one to five years. As presented in Table 7, ε_1 values are very small for all investment horizons and for both hedge fund styles, which implies that after accounting for the time variation and nonlinearities in the conditional return distributions, the directional hedge funds remain more dominant than the U.S. equity and bond markets. Comparing the results from the unconditional distributions in Table 3 with the results from the conditional distributions in Table 7 clearly indicates that the relative strength of the long/short equity hedge and emerging markets strategies becomes stronger when investors form their ex ante expectations with the conditional skewed fat-tailed distributions as opposed to forming unconditional expectations.

7. Conclusion

Because the return distribution of hedge fund portfolios as well as the distribution of equity and bond returns exhibit significant departures from normality, the classical selection rules do not provide an appropriate framework to explain investors' preferences

Table 7 Comparing Time-Varying Conditional Distributions

Long/short equity hedge					
	1-year	2-year	3-year	4-year	5-year
S&P 500	0.0006	0.0000	0.0000	0.0000	0.0000
One-month T-bill	0.0001	0.0000	0.0000	0.0000	0.0000
10-year T-bond	0.0294	0.0055	0.0088	0.0049	0.0000
Emerging markets					
S&P 500	0.0233	0.0000	0.0000	0.0000	0.0000
One-month T-bill	0.0130	0.0000	0.0000	0.0000	0.0000
10-year T-bond	0.0588	0.0000	0.0000	0.0000	0.0000

Notes. This table compares the relative performance of the long/short equity hedge and emerging markets strategies with the U.S. equity and bond markets. The parameters of the conditional SGED are estimated for the long/short equity hedge and emerging markets strategies, and the S&P 500 index, one-month and 10-year Treasury returns for the sample period January 1994–December 2011. Once the time-varying conditional distributions are obtained, the relative strength of the hedge fund market is determined based on the AFSD criterion. The table presents the empirical estimates of ε_1 for the conditional return distributions.

among these alternative investment choices. Different from the existing literature, we examine the practice of investing in hedge funds by using the ASD approach of Leshno and Levy (2002) that does not require a parametric specification of investors' preferences and does not make any assumptions about asset returns. The results indicate that popular hedge fund strategies (long/short equity hedge and emerging markets) outperform the U.S. equity market. However, the remaining nine hedge fund strategies considered in this paper do not generate superior performance over the S&P 500 index.

We also investigate the almost stochastic dominance of hedge fund and bond markets. The relative performance of hedge fund portfolios with respect to the one-month Treasury bills is found to be sensitive to the choice of investment horizon. For the short-term investment horizon of one-year, the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies outperform the one-month T-bills based on the AFSD. The remaining seven hedge fund strategies do not outperform the short-term Treasury security at the one-year horizon. For longer investment horizons, all hedge fund strategies (with the only exception of short bias) perform better than the one-month Treasury bill in the sense of AFSD. Finally, we compare the performance of hedge fund portfolios with the 10-year Treasury bonds for one- to five-year investment horizons. For all investment horizons, the global macro, managed futures, multistrategy, and long/short equity hedge fund strategies outperform the 10-year Treasury bond based on the AFSD. For the one-year investment horizon, in addition to the aforementioned four strategies, the equity market neutral and fixed income

arbitrage strategies also beat the long-term Treasury market. For four- and five-year investment horizons, most hedge fund strategies (with the exception of short bias, fund of funds, fixed income, convertible arbitrage, and equity market neutral) perform better than the 10-year Treasury security in the sense of AFSD.

In addition to the utility-based nonparametric approach of Leshno and Levy (2002), we use the utility-based parametric measure of Goetzmann et al. (2007). The MPPM of Goetzmann et al. (2007) produces very similar findings; the directional funds with dynamic trading strategies are ranked among the best performers, whereas the least directional funds with inactive trading strategies are ranked among the worst performers. However, the contradictory and unstable results are obtained from the traditional performance measures (mean-variance rule, Sharpe and Treynor ratios, and alphas). We also use measures of modified Sharpe ratio that replace standard deviation with the *VaR*, the *MDD*, and the *LPM*₂. The results indicate that the mean-*VaR* and Sortino ratios, replacing standard deviation with *VaR* and *LPM*₂, respectively, do not generate a robust, consistent ranking among hedge fund strategies. However, the results from the Calmar ratio replacing standard deviation with *MDD* are similar to those obtained from the ASD measure and MPPM. Overall, we find that the downside risk-adjusted performance measures produce conflicting evidence for the relative performance of hedge fund strategies. Hence, the ASD measure and MPPM are more appropriate when assessing the performance of hedge fund strategies.

Acknowledgments

The authors thank department editor Wei Jiang, the anonymous associate editor, and two referees for their extremely helpful comments and suggestions. The authors also benefited from discussions with George Aragon, Haim Levy, Bing Liang, Tarun Ramadorai, Leola Ross, and the seminar participants at the City University of New York and Georgetown University. Turan Bali thanks the Research Foundation of the McDonough School of Business, Georgetown University. Stephen Brown thanks the Stern School of Business, New York University. Ozgur Demirtas thanks FP7-PEOPLE-2012-CIG: Marie-Curie Action: "Career Integration Grants" for making this project possible. All errors remain the authors' responsibility.

Appendix. Simulation Analysis for the Critical Values

As discussed in §2.2, the critical values of ε_1 and ε_2 are obtained by running a series of experiments with sufficiently large subjects and would require that 100% of the subjects select high-return assets (*H*) over low-return assets (*L*) by asking them their highest required cash flows for *H* at a given state. This helps determine the smallest allowed ε_1^* and ε_2^* across all subjects, and in turn the

critical values are obtained. Levy et al. (2010) conducted experiments based on 400 subjects' choices by defining the economically relevant set of preferences and the corresponding almost decision rules that avoid the paradoxical results. The results of these experiments are very robust and are almost unaffected by the magnitude of the asset returns as well as the asset classes under consideration. Hence, we argue that the critical values used in our empirical analysis ($\varepsilon_1^* = 5.9\%$, $\varepsilon_2^* = 3.2\%$) are robust indicators of almost all investors' choice of a portfolio with favorable distributional characteristics.

However, one may still be concerned about the sensitivity of the critical values to the aforementioned experimental studies. To relieve any potential concerns, we conduct several bootstrapping and randomization analyses. In this section, we present results from three distinct simulations to determine the statistical inference of ε_1^* and ε_2^* . Specifically, these bootstrapping and randomization analyses generate the p -values corresponding to the critical values used in the paper.

We first rely on the randomization technique. We use the monthly S&P 500 index returns for the sample period from January 1926 to December 2010 and generate the distribution of ε_1 and ε_2 in repeated samples. In each run of the simulation, we pick (with replacement) two series of 1,000 monthly return observations from the S&P 500 index. We then compute values of ε_1 and ε_2 for those two series and record them. This exercise is repeated 1,000 times, which generates 1,000 ε_1 and ε_2 values. Because, in each run of the randomization, the repeated samples are drawn from the same empirical distribution of S&P 500 returns, the null hypothesis is that the two return series do not dominate each other, i.e., the computed ε_1 and ε_2 values are not expected to be lower than $\varepsilon_1^* = 5.9\%$ and $\varepsilon_2^* = 3.2\%$, respectively.

Panel A of Table A.1 presents the statistics for the distribution of ε_1 and ε_2 obtained from the randomization analyses. Specifically, the minimum, maximum, mean, standard deviation, and the 1, 5, 10, 25, 50, 75, 90, 95, and 99 percentiles of the generated ε_1 and ε_2 values are reported.

We determine with what probability the computed ε_1 and ε_2 values from randomization are lower than the critical values when the underlying distribution is in fact the same. As

expected, the mean and the median of ε_1 are close to 0.5, indicating that, on average, the cumulative distribution of one randomized return series is above the cumulative distribution of the other randomized series 50% of the time. As expected, the computed ε_1 and ε_2 are also bounded between 0 and 1. Most importantly, the left tail of the distribution shows the number of simulations (out of 1,000 runs) that generate ε_1 values that are lower than the critical value used in the paper. As shown in panel A of Table A.1, even the 5th percentile of the distribution of ε_1 is 6.25%, which is greater than $\varepsilon_1^* = 5.9\%$, which implies that the estimated ε_1 values reported in our tables have p -values lower than 5%. Similarly, the second row of panel A shows that the p -value of ε_2 is also slightly lower than 5%.

In our second set of analyses, we rely on a bootstrapping methodology. During the sample period January 1926–December 2010, the S&P 500 index has a mean of 0.61% per month and a standard deviation of 5.54% per month. Simulated return series are generated based on the assumption of normality. Specifically, in each run of the bootstrap, two return series of 1,000 monthly observations are drawn from a normal distribution with the same mean and standard deviation of the S&P 500 index. Then, ε_1 and ε_2 values are computed for each run. The analysis is repeated 1,000 times, and the distribution of ε_1 and ε_2 are generated as described earlier.

The results reported in panel B of Table A.1 turn out to be very similar to our earlier findings in panel A. The 5th percentile of the distribution of ε_1 is 6.03%, which is greater than $\varepsilon_1^* = 5.9\%$, implying that the estimated ε_1 values reported in our tables have p -values lower than 5%. Similarly, the second row of panel B shows that the p -value of ε_2 is also slightly lower than 5%.

Finally, in unreported results, we also conduct an analysis in which the two time series are drawn from a normal distribution with similar (but not identical) return moments. Specifically, one return series is drawn from a normal distribution with the mean and standard deviation of the S&P 500 index return (0.61% and 5.54% per month, respectively), and the other return series is drawn from a normal distribution with a slightly higher mean and standard deviation (0.65% and 6.00% per month, respectively). This procedure is repeated 1,000 times, and it is concluded that the

Table A.1 Summary Statistics for the Distribution of ε_1 and ε_2

Panel A: Distribution of ε_1 and ε_2 from randomization												
Min	1%	5%	10%	25%	50%	75%	90%	95%	99%	Max	Mean	Std dev
ε_1												
0	0.0017	0.0625	0.1125	0.2708	0.5294	0.7744	0.9190	0.9629	0.9936	1	0.5214	0.2913
ε_2												
0	0	0.0387	0.0575	0.1732	0.4509	0.7481	0.8864	0.9324	0.9686	0.9846	0.4611	0.3098
Panel B: Distribution of ε_1 and ε_2 from bootstrapping												
ε_1												
0.0082	0.0158	0.0603	0.0925	0.2051	0.5124	0.7993	0.9120	0.9534	0.9838	0.9948	0.5014	0.3123
ε_2												
0	0	0.0352	0.0513	0.1696	0.4402	0.7436	0.8863	0.9286	0.9653	0.9829	0.4516	0.3106

distribution of ε_1 and ε_2 provide p -values that are smaller than 5%.

Doing the statistical inference correctly is one of the main difficulties associated with the stochastic dominance rules based on the empirical distribution of returns. Consistent results obtained from the bootstrapping and randomization analyses provide evidence that the critical values ($\varepsilon_1^* = 5.9\%$, $\varepsilon_2^* = 3.2\%$) used in the paper are robust indicators of almost all investors' choice of a portfolio with favorable distributional characteristics.

References

- Ackermann C, McEnally R, Ravenscraft D (1999) The performance of hedge funds: Risk, return, and incentives. *J. Finance* 54(3):833–874.
- Agarwal V, Naik NY (2000) Multi-period performance persistence analysis of hedge funds. *J. Financial Quant. Anal.* 35(3):327–342.
- Agarwal V, Fos V, Jiang W (2013a) Inferring reporting-related biases in hedge fund databases from hedge fund equity holdings. *Management Sci.* 59(6):1271–1289.
- Agarwal V, Jiang W, Tang Y, Yang B (2013b) Uncovering hedge fund skill from the portfolio holdings they hide. *J. Finance* 68(2):739–783.
- Aggarwal RK, Jorion P (2010) The performance of emerging hedge funds and managers. *J. Financial Econom.* 96(2):238–256.
- Aragon G (2007) Share restrictions and asset pricing: Evidence from the hedge fund industry. *J. Financial Econom.* 83(1):33–58.
- Aragon G, Martin JS (2012) A unique view of hedge fund derivatives usage: Safeguard or speculation. *J. Financial Econom.* 105(2):436–456.
- Arzac ER, Bawa VS (1977) Portfolio choice and equilibrium in capital markets with safety-first investors. *J. Financial Econom.* 4(3):277–288.
- Asness CS, Krail R, Liew JM (2001) Do hedge funds hedge? *J. Portfolio Management* 28(1):6–19.
- Bali TG, Theodossiou P (2008) Risk measurement performance of alternative distribution functions. *J. Risk Insurance* 75(2):411–437.
- Bali TG, Brown SJ, Caglayan MO (2011) Do hedge funds' exposures to risk factors predict their future returns? *J. Financial Econom.* 101(1):36–68.
- Bali TG, Brown SJ, Caglayan MO (2012) Systematic risk and the cross section of hedge fund returns. *J. Financial Econom.* 106(1):114–131.
- Bali TG, Gokcan S, Liang B (2007) Value at risk and the cross-section of hedge fund returns. *J. Banking Finance* 31(4):1135–1166.
- Bali TG, Demirtas KO, Levy H, Wolf A (2009) Bonds versus stocks: Investors' age and risk taking. *J. Monetary Econom.* 56(6):817–830.
- Baumol WJ (1963) An expected gain-confidence limit criterion for portfolio selection. *Management Sci.* 10(1):174–182.
- Bollerslev T (1986) Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* 31(3):307–327.
- Brown SJ, Goetzmann WN, Ibbotson RG (1999) Offshore hedge funds: Survival and performance 1989–95. *J. Bus.* 72(1):91–117.
- Brown SJ, Gregoriou G, Pascalau R (2012) Diversification in funds of hedge funds: Is it possible to overdiversify? *Rev. Asset Pricing Stud.* 2(1):89–110.
- Brown SJ, Goetzmann WN, Ibbotson RG, Ross SA (1992) Survivorship bias in performance studies. *Rev. Financial Stud.* 5(4):553–580.
- Chen Y, Liang B (2007) Do market-timing funds time the market? *J. Financial Quant. Anal.* 42(4):827–856.
- Edwards FR, Caglayan MO (2001) Hedge fund performance and manager skill. *J. Futures Markets* 21(11):1003–1028.
- Fama EF, French KR (1993) Common risk factors in the returns of stocks and bonds. *J. Financial Econom.* 33(1):3–56.
- Fung WH, Hsieh DA (2000) Performance characteristics of hedge funds and CTA funds: Natural versus spurious biases. *J. Financial Quant. Anal.* 35(3):291–307.
- Fung WH, Hsieh DA (2001) The risk in hedge fund strategies: Theory and evidence from trend followers. *Rev. Financial Stud.* 14(2):313–341.
- Fung WH, Hsieh DA (2004) Hedge fund benchmarks: A risk-based approach. *Financial Analysts J.* 60(5):65–80.
- Fung WH, Hsieh DA, Naik NY, Ramadorai T (2008) Hedge funds: Performance, risk, and capital formation. *J. Finance* 63(4):1777–1803.
- Goetzmann WN, Ingersoll JE, Spiegel M, Welch I (2007) Portfolio performance manipulation and manipulation-proof performance measures. *Rev. Financial Stud.* 20(5):1503–1546.
- Kosowski R, Naik NY, Teo M (2007) Do hedge funds deliver alpha? A Bayesian and bootstrap analysis. *J. Financial Econom.* 84(1):229–264.
- Leshno M, Levy H (2002) Preferred by “all” and preferred by “most” decision makers: Almost stochastic dominance. *Management Sci.* 48(8):1074–1085.
- Levy H, Sarnat M (1972) Safety first: An expected utility principle. *J. Financial Quant. Anal.* 7(3):1829–1834.
- Levy H, Leshno M, Leibovitch B (2010) Economically relevant preferences for all observed epsilon. *Ann. Oper. Res.* 176(1):153–178.
- Liang B (1999) On the performance of hedge funds. *Financial Analysts J.* 55(4):72–85.
- Liang B (2000) Hedge funds: The living and the dead. *J. Financial Quant. Anal.* 35(3):309–326.
- Liang B (2003) Hedge fund returns: Auditing and accuracy. *J. Portfolio Management* 29(3):111–122.
- Malkiel BG, Saha A (2005) Hedge funds: Risk and return. *Financial Analysts J.* 61(6):80–88.
- Markowitz H (1952) Portfolio selection. *J. Finance* 7(1):77–91.
- Markowitz H (1959) *Portfolio Selection: Efficient Diversification of Investments* (Wiley, New York).
- Mitchell M, Pulvino T (2001) Characteristics of risk and return in risk arbitrage. *J. Finance* 56(6):2135–2175.
- Patton AJ (2009) Are “market neutral” hedge funds really market neutral? *Rev. Financial Stud.* 22(7):2495–2530.
- Roy AD (1952) Safety first and the holding of assets. *Econometrica* 20(3):431–449.
- Scott RC, Horvath PA (1980) On the direction of preference for moments of higher order than the variance. *J. Finance* 35(4):915–919.