Performance Evaluation with High Moments and Disaster Risk[1]

Ohad Kadan[2] Fang Liu[3]

February 2014

[1]We thank an anonymous referee, Phil Dybvig, Sergiu Hart, Isaac Kleshchelski, Asaf Manela, Mark Rubinstein, and Shlomo Yitzhaki as well as seminar participants at the University of Arizona, Interdisciplinary Center, Hertzlia, the Institute of Financial Studies at Southwestern University of Finance and Economics, China, and Washington University in St. Louis for helpful comments and suggestions.

[2]Olin Business School, Washington University in St. Louis. E-mail: kadan@wustl.edu.

[3]Olin Business School, Washington University in St. Louis. E-mail: fliu23@wustl.edu.

Abstract

Traditional performance evaluation measures do not account for tail events and rare disasters. To address this issue, we reinterpret the riskiness measures of Aumann and Serrano (*Journal of Political Economy*, 2008) and Foster and Hart (*Journal of Political Economy,* 2009) as performance indices. We derive the moment properties of these indices and their sensitivity to rare disasters and show that they are consistent with the asset pricing literature. As applications, we show that "anomalous" investment strategies such as "momentum" or investment in private equity lose much of their glamour when accounting for high moments and rare events. Furthermore, using the indices to select mutual funds results in desirable high-moment properties out of sample.

1. Introduction

Tail risk and rare disasters have been central to the recent meltdown in financial markets. Indeed, markets were hit by catastrophic events whose exante probabilities were considered negligible. Traditional performance evaluation measures (such as the Sharpe ratio) typically rely on the first two distribution moments, thereby underestimating the effects of rare disasters. Indeed, low distribution moments hardly account for rare and catastrophic events, since their large negative effect is multiplied by a very small probability. By contrast, when one considers high distribution moments, an extremely negative but rare outcome is raised to a high power, making its effect on the moment substantial regardless of the small probability associated with it.

High distribution moments have received notable attention in the asset pricing literature. In particular, a large body of work in asset pricing suggests that investors favor right skewness (e.g., Rubinstein, 1973; Kraus and Litzenberger, 1976; Jean, 1971; Kane, 1982; Harvey and Siddique, 2000), but are averse to tail-risk and rare disasters (e.g., Barro, 2006, 2009; Gabaix, 2008, 2012; Gourio, 2012; Chen, Joslin, and Tran, 2012; Wachter, 2013). It is thus desirable that normative performance evaluation measures reflect these preferences.

In this paper we study two such performance indices relying on a simple reinterpretation of the novel riskiness measures proposed by Aumann and Serrano (2008) and Foster and Hart (2009) (hereafter AS and FH, respectively).[1] We investigate the moment properties of these indices and establish that they reflect *all* distribution moments in a manner consistent with economic intuition and with the asset pricing literature. We also discuss the way these two indices reflect disaster risk. We then apply these indices to popular investment strategies and to well-known anomalies, show their practical usefulness in selecting mutual funds, and demonstrate the pitfalls associated with ignoring high moments and rare disasters in performance evaluation.

Our starting point is that investors are risk-averse and choose their investments

[1] Aumann and Serrano (2008) offer a set of axioms characterizing the AS riskiness measure. An axiomatization of the FH measure is offered separately in Foster and Hart (2013).

by maximizing expected utility. The best possible way to rank investments in this setup is known to be Second Order Stochastic Dominance (SOSD) (see Hadar and Russell, 1969; Hanoch and Levy, 1969; Rothschild and Stiglitz, 1970), according to which one investment dominates another if all risk-averse investors prefer the former to the latter. The problem with SOSD is that it only imposes a partial order on investments. Namely, some pairs of investments cannot be ranked using SOSD.

Based on our discussion thus far, a desirable performance evaluation index should satisfy the following four requirements: (i) Impose a complete order on investments, namely, any two investments can be compared; (ii) Depend on the distribution of outcomes only. That is, the form of the utility function is not needed to calculate the performance index; (iii) Coincide with SOSD, whenever SOSD can be applied. Namely, if all risk-averse investors prefer one investment to the other, then the performance index ranks the investments accordingly; and (iv) Account for high distribution moments in a manner consistent with the asset pricing literature. That is, the index is increasing in mean and skewness and decreasing in variance and tail-risk of the investment.

The Sharpe ratio, which is probably the most popular performance evaluation measure, satisfies (i) and (ii), but clearly fails (iv). Interestingly, it also fails (iii). Indeed, it is fairly easy to find examples in which all risk-averse investors prefer one investment to the other and yet the Sharpe ratio ranks the investments in the wrong order (see Section 2 for examples). In Appendix B we review several other popular performance evaluation measures and discuss the extent to which they satisfy these four requirements.

To understand the fundamental insights in AS and FH it is useful to follow the approach presented in Hart (2011), who offers a unified framework for the two. The key for the new indices is to use the investor's initial wealth as a benchmark for her investment decisions. That is, instead of comparing the expected utility of two investments, we compare the expected utility of each investment separately to the status quo, and ask which one of the two investments is uniformly rejected more

often. If each time that investment g is uniformly rejected we have that investment g' is also uniformly rejected, then g is deemed more attractive than g' (i.e., g has better performance than g').[2] That is, g is more attractive than g' if g is rejected "less often" than g' in some uniform manner when compared to the status quo.

The term "uniform rejection" can take two different meanings. First is "wealth-uniform rejection" in which for a given utility function, an investor rejects the investment relative to the status quo for all wealth levels. Second is "utility-uniform rejection" in which for a given wealth level, all utility functions reject the investment relative to the status quo. The former approach to uniform rejection leads to the AS performance index, while the latter leads to the FH performance index.

As shown in AS, FH, and Hart (2011), the two approaches yield two rankings of investments, each of which can be represented by a positive performance index that possesses an intuitive economic interpretation. Both indices satisfy requirements (i)–(iii) above. Moreover, they can be easily calculated from the distribution of the investment by solving an intuitive implicit equation. The only difference between our interpretation and the interpretations in AS and FH is that they choose to consider the riskiness of the investment, deeming one investment *"more risky"* than another if it is uniformly rejected *more* often relative to the status quo. We choose to focus on the flip side of the argument, viewing one investment as *"more attractive"* than another if it is uniformly rejected *less* often relative to the status quo. Roughly speaking, we view an investment as "attractive" if risk-averse investors show little aversion to this investment when compared to the status quo, in a uniform manner.

The first thing we do in this paper is to extend the AS and FH indices to a multi-period setting. We show that the AS and FH results can readily be considered in such a setting, and that if gambles are identically distributed in each period, then the multi-period performance indices coincide with the single-period indices.

We then turn to studying how the AS and FH performance indices are affected

[2] The term "investment" here simply refers to a random variable which can be described by the probability distribution over outcomes. We often use the term "gamble," which is the one used in AS and FH, instead. We use the letter g as a generic notation for such investments (or gambles).

by the moments of the investments being evaluated. We establish that both the AS and FH indices reflect *all* the distribution moments (raw and central). Moreover, these performance indices are increasing in all odd moments and decreasing in all even moments. Consequently, the two indices satisfy requirement (iv) above.

Next, we ask whether the sensitivity of the performance indices to the moments is monotonically decreasing in the order of the moment. Namely, do high distribution moments necessarily have a smaller effect on performance than low distribution moments? We establish that there is no such monotone relation. In particular, the performance indices can be either more or less sensitive to higher moments. Thus, high moments can have a material effect on performance, and should not be neglected.

We then turn to exploring how the performance indices are affected by rare disasters, modeled as extremely negative outcomes associated with vanishing probabilities. First, note that such outcomes tend to make the distribution left skewed (more negative third moment) and fat-tailed (higher fourth moment). Thus, given requirement (iv), both performance indices are adversely affected by rare disasters. However, we show that the FH index is much more sensitive to rare disasters than the AS index.

When making decisions, investors often face exogenous and unavoidable risks such as macroeconomic shocks and shocks to labor income. This kind of uncertainty is often termed "background risk." In our final theoretical analysis we study how such background risk affects the AS and FH indices. We consider two approaches to modeling background risk. The first is additive, where exposure to background risk is modeled by adding a random shock to the investor's initial wealth. The second approach is multiplicative, where the final wealth of the investor is multiplied by a random shock. We show that the AS index lends itself naturally to the additive approach, while the FH index fits well into the multiplicative approach. Furthermore, we find that if modeled this way, background risk does not affect the AS and FH performance indices, and so essentially it could be ignored.

We next turn to exploring the practical implications of the two performance indices. To this end, we show that the two indices lend themselves naturally to es-

timation using the Generalized Method of Moments (GMM) (see Hansen, 1982). This approach allows us to test hypotheses regarding the attractiveness of different investment strategies in the underlying population of returns.

We first use the performance indices to evaluate the most prominent and widely studied investment anomalies: the size anomaly, the value anomaly, and the momentum anomaly. We compare these investment strategies to each other and to a naive "buy and hold" strategy of investing in the market. We do this by examining the performance of the four Fama-French factors (Fama and French, 1993; Carhart, 1997). Our most interesting finding here is that the momentum strategy, often considered the most serious deviation from market efficiency (Fama and French, 1996), is no longer attractive when accounting for high moments. Momentum returns are extremely left skewed [as originally pointed out by Harvey and Siddique (2000)] and fat-tailed, and they exhibit extreme negative events, which fall under our definition of "rare disasters." These high-moment properties outweigh the higher average return obtained from following momentum. In particular, our estimates of the AS and FH performance measures show that momentum does not have better performance than a "buy and hold" investment in the market. Moreover, we find that momentum is dominated by the value anomaly, and it remains dominated even when combined with other anomalies.

In our next application we compare the performance of private investments to public equity. Moskowitz and Vissing-Jorgensen (2002) find that the returns to private equity are not higher than those of public equity. They view this result as puzzling since private equity investments expose investors to a high level of idiosyncratic risk. Moskowitz and Vissing-Jorgensen note that private equity returns are right skewed and conjecture that preference for skewness may be one reason for the tendency of individuals to invest in private equity. The indices studied in this paper are useful for evaluating this statement since they take into account all distribution moments (skewness among them). Thus, we follow Moskowitz and Vissing-Jorgensen (2002) and compare the returns of public investments to those of private investments

obtained from the 2004 Survey of Consumer Finance (SCF). We find that the average return on private equity conditional on survival is about 35 times larger than that of public equity. Moreover, private equity returns are indeed very right skewed. However, private equity returns are also extremely more volatile and fat-tailed than the returns on public equity. The question is then whether the superior first and third moments of private equity outweigh its inferior second and fourth moments. Our estimates of the two indices suggest that this is not the case. Both the AS and FH indices are significantly higher for public investments. Thus, based on our estimates, the "private equity premium puzzle" suggested by Moskowitz and Vissing-Jorgensen (2002) still stands, and is not resolved by high-moment properties.

In the next application we compare the performance of actively managed equity funds to that of index funds. The question is whether the returns for actively managed funds exceed those of passive funds controlling for risk. Given that investors care about all moments of the return distribution, we extend standard analyses to account for those moments using the new performance measures. We find that the moments of the two management strategies are not materially different. Moreover, our estimates show that the performance indices of active vs. passive mutual funds (after accounting for fees) are not significantly different. Thus, the new performance indices reinforce the view that active management does not improve investment performance (even when considering high distribution moments).

In our final analysis we take the two performance indices one step farther. Rather than just examining the performance of investment strategies, we use the indices to select among actively managed mutual funds, and examine the performance resulting from such an investment strategy. If the high-moment properties of investment portfolios are persistent, then we expect portfolios sorted on the AS and FH measures to exhibit superior performance out of sample.

To test this, in each month during our sample period of 1967–2009, we rank all actively managed equity mutual funds based on their historical AS and FH indices. We then obtain two portfolios of "selected" mutual funds by equal-weighting the

funds in the top decile for each index. We compare these two portfolios to the market portfolio and to a portfolio selected based on the Sharpe ratio. We find that moments generated by the AS and FH indices are significantly more appealing than those generated by the Sharpe ratio and are also often more attractive than those of the market portfolio. In particular, portfolios of mutual funds based on the AS and FH indices have lower variance, less negative skewness, and lower tail-risk than the market or the Sharpe ratio-based portfolios. Reflecting these observations, both the AS and FH indices are higher for portfolios which select mutual funds based on these two indices. This suggests that the two indices may be useful not only for evaluating investments but also for selecting investments that have desirable moment properties.

Overall, our empirical results demonstrate the importance of high moments in performance evaluation. What looks like an attractive investment strategy when focusing on the first two moments, can easily become less appealing when considering higher moments and disaster risk. The paper thus contributes to the performance evaluation literature and to our understanding of the abnormal returns associated with different trading strategies. It also contributes to the growing literature which applies the AS and FH measures. For example, Bakshi, Chabi-Yo, and Gao (2011) use the Aumann and Serrano (2008) riskiness measure to study how changes in riskiness over time affect the equity, value, size, and momentum premiums.

We proceed as follows. Section 2 presents motivating examples. Section 3 introduces the two performance indices. In Section 4 we derive properties of the two indices. Section 5 discusses practical applications of the indices to different investment strategies. Section 6 studies the behavior of mutual fund portfolios selected using the two indices. We conclude in Section 7. All proofs are in Appendix A.

2. Motivating examples

Before discussing the new performance indices, we consider two examples highlighting distributional features that fail to be captured by the Sharpe ratio. We will show later that the new performance indices successfully incorporate these features.

The first example, as shown in Table 1, involves the comparison of two gambles. Gamble g_1 looks like a relatively safe bet. However, it assigns a very small probability to a rare but disastrous event of losing 10. By contrast, g_2 is more volatile than g_1. Yet, the distribution of g_2 lies (weakly) to the right of that of g_1. Hence, g_2 first-order stochastically dominates g_1. That is, all investors with increasing utility prefer g_2 to g_1 (regardless of risk attitude). But, the Sharpe ratio of g_1 is higher than that of g_2. This reflects the fact that the variance of g_1 is very low. In this case, the Sharpe ratio fails to capture the preference of any reasonable investor.

The problem with the Sharpe ratio is tied closely to the high moments of the gambles. Notice that both the mean and the variance of g_1 are only mildly affected by the rare disaster. However, higher moments would be more strongly affected by this event. Indeed, by the way that higher moments are calculated, a disastrous outcome is raised to a higher and higher power, while the probability associated with it does not change. As a result, for sufficiently high moments, disastrous outcomes dominate the low probability assigned to them, and hence have a material effect on the moment itself. Investors maximizing expected utility care about all moments of the distribution. Hence, disastrous but rare events such as in g_1 may have a material effect on their preferences. In the example above we have calculated the third and fourth central moments of g_1 and g_2 for illustration (denoted by m_3 and m_4). It can be seen that the third moment of g_1 is negative while the third moment of g_2 is positive, reflecting the left skewness of g_1 compared to g_2. And, the fourth moment of g_1 is larger than that of g_2, reflecting the tail-risk associated with g_1. These high moments are incorporated into the decisions of expected utility maximizers. Our view is that they should also be incorporated into performance evaluation indices.

One would wonder whether the failure of the Sharpe ratio in the previous example is driven by the existence of the rare disaster. The second example (presented in Table 2) suggests that it is not the case. In this example, we replace the disastrous event of g_1 by a mild loss of -1. This does not change the fact that gamble g_2 first-order stochastically dominates g_1. Once again, the Sharpe ratio dramatically favors the

wrong gamble as it is almost 11 times higher for g_1 than for g_2. We conclude that the Sharpe ratio may fail to capture the preferences of any reasonable investor even in the absence of a rare disaster. The calculation of higher moments suggests that despite the fact that g_2 has a higher fourth central moment (m_4), it is also right skewed (reflected by a positive m_3) as opposed to the left skewness of g_1. Thus, the high fourth moment of g_2 is attributed to the right tail. It seems reasonable that these high-moment properties should also be accounted for in performance evaluation.

3. The performance indices

In this section we first review and reinterpret relevant results in Aumann and Serrano (2008) and Foster and Hart (2009). To do so, we follow the unified approach presented in Hart (2011). Then, we extend these results to a multi-period setting.

3.1. One-period gambles

An investment can be modeled as a random variable, which we generically denote by g. We assume that all moments of g are well defined. Furthermore, we assume that g has positive expectation and that it admits some negative values with positive probability. We often refer to g as a "gamble" and denote the set of all such gambles by $\mathcal{G}$. For the FH measure only, we also require that g be bounded from below.

We assume that investors have von Neumann-Morgenstern utility functions over wealth denoted by $u(\cdot)$, which are differentiable as many times as needed. We assume further that $u' > 0$ and $u'' < 0$, reflecting that investors like more wealth over less, and are strictly risk-averse. Furthermore, we restrict attention to utility functions u satisfying the following three conditions: (i) Decreasing absolute risk aversion (DARA), i.e., $-\frac{u''(w)}{u'(w)}$ is weakly decreasing; (ii) Increasing relative risk aversion (IRRA), i.e., $-w\frac{u''(w)}{u'(w)}$ is weakly increasing; and (iii) $\lim_{w\downarrow 0} u(w) = -\infty$. We denote the class of all such utility functions by $\mathcal{U}^*$, and note that this class includes, for example, all constant relative risk aversion (CRRA) utility functions of the form $u(w) = \frac{w^{1-\gamma}}{1-\gamma}$ with $\gamma \geq 1$, as well as utility functions that are constant absolute risk aversion (CARA)

from a sufficiently high wealth level on.

Let w_0 denote the initial wealth of an investor, to which we refer as her "status quo."

Definition 1. *Say that an investor with utility u and initial wealth w_0 rejects a gamble g if $\mathbf{E}\left[u\left(w_0+g\right)\right] \leq u\left(w_0\right)$, and accepts a gamble g if $\mathbf{E}\left[u\left(w_0+g\right)\right] > u\left(w_0\right)$.*

That is, an investor rejects a gamble whenever her status quo yields her a weakly higher expected utility. The following two definitions are needed to describe the Aumann-Serrano performance index.

Definition 2. *Say that a gamble g is wealth-uniformly rejected by an investor with utility function u, if u rejects g at all initial wealth levels w_0.*

Intuitively, an investor wealth-uniformly rejects a gamble g, if she prefers the status quo to g, regardless of her wealth level.

Definition 3. *Say that a gamble g wealth-uniformly dominates gamble g' if whenever g is wealth-uniformly rejected by a utility function u, g' is also wealth-uniformly rejected by u.*

Namely, g wealth-uniformly dominates g' if whenever an investor with utility function u prefers the status quo to g for all wealth levels, she also prefers the status quo to g' for all wealth levels. In other words, g is preferred to g', if g' is "more often" wealth-uniformly rejected than g is.

Proposition 1. *(Aumann and Serrano, 2008; Hart, 2011). Wealth-uniform dominance induces a complete order on $\mathcal{G}$ that extends SOSD. This order can be represented by a performance index $P^{AS}(g)$ assigned to any gamble $g \in \mathcal{G}$, which is given by the unique positive solution to the implicit equation*

$$\mathbf{E}\left[\exp\left(-P^{AS}(g)g\right)\right] = 1. \tag{1}$$

That is, for any two gambles g and g', g wealth-uniformly dominates g' if and only if $P^{AS}(g) \geq P^{AS}(g')$.

To gain intuition for the performance index P^{AS}, it is useful to rewrite (1) as

$$\mathbf{E}\left[-\exp\left(-P^{AS}(g)(w_0+g)\right)\right] = -\exp\left(-P^{AS}(g)\,w_0\right), \tag{2}$$

for some initial wealth w_0. Note that (1) and (2) are equivalent regardless of w_0. Thus, a useful interpretation is that $P^{AS}(g)$ is the level of absolute risk aversion that makes an investor with CARA utility indifferent between taking g and the status quo, regardless of the initial wealth w_0. Put differently, an investor with CARA utility $u(w) = -\exp(-\lambda w)$, would accept g when $\lambda < P^{AS}(g)$ and would reject g when $\lambda \geq P^{AS}(g)$. Thus, a higher level of $P^{AS}(g)$ means that investors are "less averse" to g, in the sense that a higher level of risk aversion is needed to reject g. The key insight in Proposition 1 is that checking (1) is both necessary and sufficient for wealth-uniform dominance for all utilities in $\mathcal{U}^*$. As such, a higher level of $P^{AS}(g)$ reflects better performance for all utility functions in $\mathcal{U}^*$ in the sense that the gamble is wealth-uniformly rejected by a smaller set of utility functions.

To understand the source of this insight it is useful to consider a situation in which there is an upper bound on initial wealth levels denoted by $\bar{w}$. Then, since $\mathcal{U}^*$ only includes DARA utility functions, a gamble g is wealth-uniformly rejected by u if and only if it is rejected by u at $\bar{w}$. Let $A_u(\bar{w})$ denote the absolute risk aversion of u at initial wealth $\bar{w}$. Now, $P^{AS}(g)$ is defined as the level of absolute risk aversion that renders a CARA utility investor indifferent between $w_0 + g$ and w_0 for any w_0, in particular, for $w_0 = \bar{w}$. Thus, g is wealth-uniformly rejected by u if and only if $A_u(\bar{w}) \geq P^{AS}(g)$. Consequently, $P^{AS}(g_1) \geq P^{AS}(g_2)$ if and only if any time that g_1 is wealth-uniformly rejected, also g_2 is wealth-uniformly rejected.

The next definitions are needed to describe the Foster-Hart performance index.

Definition 4. *Say that a gamble g is utility-uniformly rejected at an initial wealth level w_0 if all utility functions $u \in \mathcal{U}^*$ reject g at w_0.*

That is, a gamble g is utility-uniformly rejected at wealth level w_0, if any investor, regardless of her utility function, prefers the status quo to g at w_0.

Definition 5. *Say that a gamble g utility-uniformly dominates gamble g' if whenever g is utility-uniformly rejected at an initial wealth level w_0 , g' is also utility-uniformly rejected at w_0.*

Namely, g utility-uniformly dominates g' if whenever all investors with initial wealth level w_0 prefer the status quo to g, they also prefer the status quo to g'. Roughly, g is preferred to g', if g' is "more often" utility-uniformly rejected than g is.

Proposition 2. *(Foster and Hart, 2009; Hart, 2011). Utility-uniform dominance induces a complete order on $\mathcal{G}$ that extends SOSD. This order can be represented by a performance index $P^{FH}(g)$ assigned to any gamble $g \in \mathcal{G}$, which is given by the unique positive solution to the implicit equation*

$$\mathbf{E}\left[\log\left(1 + P^{FH}(g)g\right)\right] = 0. \tag{3}$$

That is, for any two gambles g and g', g utility-uniformly dominates g' if and only if $P^{FH}(g) \geq P^{FH}(g')$.

To gain intuition for the performance index P^{FH}, it is useful to rewrite (3) as

$$\mathbf{E}\left[\log\left(\frac{1}{P^{FH}(g)} + g\right)\right] = \log\left(\frac{1}{P^{FH}(g)}\right). \tag{4}$$

That is, $\frac{1}{P^{FH}(g)}$ can be interpreted as the level of wealth that would render an investor with log utility indifferent between taking g or staying with the status quo. A log investor with higher initial wealth than $\frac{1}{P^{FH}(g)}$ would accept g, whereas a log investor with lower initial wealth than $\frac{1}{P^{FH}(g)}$ would reject g. Thus, higher P^{FH} corresponds to better performance in the sense that g is accepted even by individuals with lower initial wealth. The key insight in Proposition 2 is that checking (3) is both necessary and sufficient for utility-uniform dominance for all initial wealth levels.

To understand this insight recall that $\mathcal{U}^*$ only includes utility functions that demonstrate IRRA and $\lim_{w\downarrow 0} u(w) = -\infty$. These properties imply that the coefficient of relative risk aversion of any utility in $\mathcal{U}^*$ must be at least 1. Thus, the utility with the lowest coefficient of relative risk aversion in $\mathcal{U}^*$ is the log utility, which

for a fixed wealth level also minimizes the coefficient of absolute risk aversion. Thus, a gamble g is utility-uniformly rejected at wealth w_0 if and only if it is rejected at w_0 by a log investor. Since $P^{FH}(g)$ is the reciprocal of the wealth level that renders a log investor indifferent between $w_0 + g$ and w_0, we have that g is utility-uniformly rejected at w_0 if and only if $w_0 \leq \frac{1}{P^{FH}(g)}$. Consequently, $P^{FH}(g) \geq P^{FH}(g')$ if and only if whenever g is utility-uniformly rejected, also g' is utility-uniformly rejected.

It is worth noting that both AS and FH present their measures as "riskiness indices" rather than "performance indices." However, this is just a matter of interpretation. Their focus is on whether investors are more reluctant to accept one gamble over another, whereas we adopt the traditional performance measurement approach in which gambles that investors are more willing to accept receive a higher score. Given this, the mapping to the original papers (Aumann and Serrano, 2008; Foster and Hart, 2009; Hart, 2011) is $P^{AS} = 1/R^{AS}$ and $P^{FH} = 1/R^{FH}$, where R^{AS} and R^{FH} are the relevant riskiness measures.

Based on the discussion thus far we conclude that the two performance indices P^{AS} and P^{FH} satisfy requirements (i)–(iii) in the Introduction. In Section 4 we study the moment properties of the indices and the way they reflect disaster risk. In particular, we establish that they also satisfy requirement (iv).

3.2. Multi-period gambles

In the context of financial investments it is natural to consider uncertain investments over time. In this section we extend the measures to a simple multi-period setting. Let T denote a finite number of periods, and consider a T-period gamble $g = \left(g^1, g^2, ...g^T\right) \in \mathcal{G}^T$, where $g^t \in \mathcal{G}$ for all $t = 1, 2, ..., T$. Consider an investor with a time separable utility function $U : \mathbb{R}_+^T \longmapsto \mathbb{R}$ that takes the form

$$U\left(w^1, w^2, ..., w^T\right) = \sum_{t=1}^{T} \rho^{t-1} u\left(w^t\right), \tag{5}$$

where $u \in \mathcal{U}^*$, $\rho \in (0, 1)$ is a discount factor, and $\left(w^1, w^2, ..., w^T\right)$ denote the wealth levels consumed by the investor in each of the T periods. The investor is endowed

with a fixed amount w_0 at the beginning of each period. To facilitate the analysis, we assume that the investor consumes her entire wealth at each period. Hence, the utility of staying with the status quo is equal to

$$U\left(w_0, w_0, ..., w_0\right) = u\left(w_0\right) \sum_{t=1}^{T} \rho^{t-1}, \tag{6}$$

and the expected utility obtained from accepting gamble g is

$$\mathbf{E}\left[U\left(w_0 + g^1, w_0 + g^2, ..., w_0 + g^T\right)\right] = \sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[u\left(w_0 + g^t\right)\right]. \tag{7}$$

The following proposition extends the P^{AS} measure to the multi-period setting.

Proposition 3. *Wealth-uniform dominance induces a complete order on $\mathcal{G}^T$. This order can be represented by a performance index $P^{AS}(g)$ assigned to any T-period gamble $g = \left(g^1, g^2, ...g^T\right) \in \mathcal{G}^T$, which is given by the unique positive solution to the implicit equation*

$$\sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[\exp\left(-P^{AS}(g) \cdot g^t\right)\right] = \sum_{t=1}^{T} \rho^{t-1}. \tag{8}$$

That is, for any two gambles g and g', g wealth-uniformly dominates g' if and only if $P^{AS}(g) \geq P^{AS}(g')$.

As in the one-period case, P^{AS} can be viewed as the risk aversion level that renders a CARA investor indifferent between taking and rejecting the T-period gamble g. Unlike in the one-period setting, here P^{AS} depends on the subjective discount factor ρ. Yet, if all g^t's are identically distributed, (8) reduces to

$$\mathbf{E}\left[\exp\left(-P^{AS}(g) g^t\right)\right] = 1, \tag{9}$$

which coincides with the one-period case, and the dependence on ρ vanishes.

Similarly, for the P^{FH} measure we have

Proposition 4. *Utility-uniform dominance induces a complete order on $\mathcal{G}^T$. This order can be represented by a performance index $P^{FH}(g)$ assigned to any T-period*

gamble $g = \left(g^1, g^2, ...g^T\right) \in \mathcal{G}^T$*, which is given by the unique positive solution to the implicit equation*

$$\sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[\log\left(1 + P^{FH}(g) g^t\right)\right] = 0. \tag{10}$$

That is, for any two gambles g *and* g'*,* g *utility-uniformly dominates* g' *if and only if* $P^{FH}(g) \geq P^{FH}(g')$.

As in the one-period case, the $P^{FH}(g)$ measure can be viewed as the reciprocal of the critical wealth level at which a log investor would be indifferent between taking and not taking gamble g. In particular, if all g^t's are identically distributed, (10) reduces to the one-period version, i.e.,

$$\mathbf{E}\left[\log\left(1 + P^{FH}(g) g^t\right)\right] = 0. \tag{11}$$

The next proposition studies the dependence of the performance measures on the subjective discount factor ρ. For convenience, we assume $T = 2$.

Proposition 5. *For any* $g = \left(g^1, g^2\right) \in \mathcal{G}^2$ *and* $P = P^{AS}$ *or* $P = P^{FH}$*,* $\frac{\partial P(g)}{\partial \rho}$ *has the same sign as* $P\left(g^2\right) - P\left(g^1\right)$.

The intuition is clear. When g_2 is a better gamble than g_1 in the sense that $P\left(g^2\right) > P\left(g^1\right)$, the two-period gamble $\left(g^1, g^2\right)$ becomes more favorable when ρ is higher, i.e., when more weight is assigned to g_2.

4. Properties of the performance indices

In this section we study the moment properties of the performance indices, their sensitivity to rare disasters, the effect of scale, leverage, and diversification on performance, and the effect of background risk. For brevity we only consider the one-period setting in this section. All of the results apply also to the multi-period setting.

4.1. Basic moment properties of the performance indices

For any gamble $g \in \mathcal{G}$, let $\mu_n(g) = \mathbf{E}[g^n]$ be the nth raw moment of g ($n \geq 1$ an integer) and let $m_n(g) = \mathbf{E}[(g - \mu_1(g))^n]$ be the nth central moment of g ($n \geq 2$ an integer). Since $g \in \mathcal{G}$, all these moments exist.

Any two gambles may differ in several of their moments. To get a basic understanding of how different moments are related to the performance indices, it is useful to consider the hypothetical exercise of changing one moment at a time while keeping all other moments fixed. For example, one can think of two investment opportunities that have identical moments except that the returns of one are more skewed than the returns of the other (higher third moment). How does this affect the performance indices of the two investments?

To see how the moments of a gamble affect its performance indices we consider first the P^{AS} index. Start by rewriting Eq. (1) as a Taylor expansion around zero:

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \left(P^{AS}(g)\right)^n \mu_n(g) = 0. \tag{12}$$

Thus, $P^{AS}(g)$ is given implicitly by the sum of a power series with coefficients proportional to the raw moments of the distribution of g. Odd moments are assigned negative weights, while even moments are assigned positive weights. A similar relation can be written with the central moments using a Taylor series around $\mu_1(g)$:

$$1 + \sum_{n=2}^{\infty} \frac{(-1)^n}{n!} \left(P^{AS}(g)\right)^n m_n(g) = \exp\left(P^{AS}(g)\,\mu_1(g)\right). \tag{13}$$

In the next proposition we use these representations to show that the P^{AS} measure is increasing in all odd moments (both raw and central) and decreasing in all even moments (both raw and central).

Proposition 6. *Consider two gambles $g, g' \in \mathcal{G}$ and let k be a positive integer.*

1. *Assume that for all $n \neq k$, $\mu_n(g) = \mu_n(g')$ but $\mu_k(g) > \mu_k(g')$. Then, $P^{AS}(g) > P^{AS}(g')$ if k is odd, while $P^{AS}(g) < P^{AS}(g')$ if k is even.*

2. *Assume that $\mu_1(g) = \mu_1(g')$ and that for all $n \neq k$, $m_n(g) = m_n(g')$ but $m_k(g) > m_k(g')$. Then, $P^{AS}(g) > P^{AS}(g')$ if k is odd, while $P^{AS}(g) < P^{AS}(g')$ if k is even.*

Next we consider the P^{FH} measure following the same approach as above. Start by rewriting Eq. (3) as a Taylor expansion around zero:

$$\sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left(P^{FH}(g)\right)^n \mu_n(g) = 0. \tag{14}$$

Notice that this Taylor expansion converges only when $-1 < P^{FH}(g)\, g \leq 1$ for all realizations of g. As before, the P^{FH} index is also given implicitly by the sum of a power series with coefficients proportional to the raw moments of the distribution of g. However, odd moments are now assigned positive weights, whereas even moments are assigned negative weights. A similar relation can be written with the central moments using a Taylor series around $\mu_1(g)$:

$$\log\left(1 + P^{FH}(g)\mu_1(g)\right) = \sum_{n=2}^{\infty} \frac{(-1)^n}{n} \left(\frac{P^{FH}(g)}{1 + P^{FH}(g)\mu_1(g)}\right)^n m_n(g). \tag{15}$$

This expansion converges for all $g \in \mathcal{G}$ such that $-1 < P^{FH}(g)\, g \leq 1 + 2P^{FH}(g)\, \mu_1(g)$.

The next proposition shows that the P^{FH} measure is also increasing (decreasing) in all odd (even) raw and central moments.

Proposition 7. *Consider two gambles $g, g' \in \mathcal{G}$ and let k be some positive integer.*

1. *Assume that for all $n \neq k$, $\mu_n(g) = \mu_n(g')$ but $\mu_k(g) > \mu_k(g')$. Suppose that $-1 < P^{FH}(g)\, g \leq 1$. Then, $P^{FH}(g) > P^{FH}(g')$ if k is odd, while $P^{FH}(g) < P^{FH}(g')$ if k is even.*

2. *Assume that $\mu_1(g) = \mu_1(g')$ and that for all $n \neq k$, $m_n(g) = m_n(g')$ but $m_k(g) > m_k(g')$. Suppose that $-1 < P^{FH}(g)\, g \leq 1 + 2P^{FH}(g)\, \mu_1(g)$. Then, $P^{FH}(g) > P^{FH}(g')$ if k is odd, while $P^{FH}(g) < P^{FH}(g')$ if k is even.*

A caveat is that unlike the P^{AS} measure, the moment properties for the P^{FH} measure only apply when the gamble is bounded from both below and above. When these conditions are not satisfied, the moment properties for the P^{FH} measure discussed in this section may not apply.

Propositions 6 and 7 tell us among other things that P^{AS} and P^{FH} are increasing in the mean and decreasing in the variance of a gamble, which is consistent with

traditional performance measures, in particular, the Sharpe ratio. In fact, it is shown in Aumann and Serrano (2008) that when a gamble g has a normal distribution, $P^{AS}(g) = 2\mu_1(g)/m_2(g)$. That is, in the normal case the P^{AS} index is proportional to a mean-to-variance ratio. For general distributions, both indices admit larger values when the third moment is large and when the fourth moment is small. Thus, requirement (iv) is satisfied by both indices.

4.2. Magnitude of the moment effects

Having established the basic moment properties, we now turn to studying the magnitude of their effects. While it has been increasingly acknowledged that higher moments play an important role in performance evaluation, standard performance indices often do not account for these aspects. For example, the widely used Sharpe ratio does not account for moments above the second, implicitly assuming that they should be assigned a negligible weight in the performance measure. It is interesting to examine whether the weight assigned to the different moments in the new performance indices is monotonically decreasing in the order of the moment.

A difficulty in examining the relative importance of the moments is that each moment is stated in a different unit of measurement. For example, suppose that the first moment (the mean) is measured in percentage points, then the second moment is measured in percentage points squared, etc. To account for this fact and allow for a "fair" comparison, we examine the magnitude effects of the "normalized moments" $\hat{\mu}_k \equiv \sqrt[k]{\mu_k}$ for $k = 1, 2, ...$ and $\hat{m}_k \equiv \sqrt[k]{m_k}$ for $k = 2, 3, ...$, since all of these have the same units of measurement as the gamble itself. For example, while m_2 is the variance of the gamble, $\hat{m}_2$ is the standard deviation. Note that moments of degree k are homogeneous of degree k, while all normalized moments are homogeneous of degree 1. Additionally, both P^{AS} and P^{FH} are homogeneous of degree -1.[3]

To gauge the influence of a moment on the performance index, we calculate the elasticity of the index with respect to normalized moments. This gives us a "unit

[3] A real function $h(\cdot)$ is homogeneous of degree k if $h(tx) = t^k h(x)$ for all x and $t > 0$.

free" estimate of the sensitivity. Since we are focusing on the magnitude of the effects rather than their directions (which we have established already), we only consider the absolute values of these elasticities (which we term "absolute elasticities"). We begin with studying the absolute elasticity of P^{AS} and P^{FH} with respect to $\hat{\mu}_k$, which we denote by η_k^{AS} and η_k^{FH}, respectively.

Consider the P^{AS} measure first. For all $g \in \mathcal{G}$, implicitly differentiating (12) yields

$$\eta_k^{AS}(g) \equiv \left| \frac{\partial P^{AS}}{\partial \hat{\mu}_k} \cdot \frac{\hat{\mu}_k}{P^{AS}(g)} \right| = \frac{\frac{1}{(k-1)!}(P^{AS}(g))^k \left|\mu_k(g)\right|}{\sum_{n=1}^{\infty} \frac{(-1)^n}{(n-1)!} \left(P^{AS}(g)\right)^n \mu_n(g)}. \tag{16}$$

Note that η_k^{AS} is homogeneous of degree 0. This follows because μ_k is homogeneous of degree k and P^{AS} is homogeneous of degree -1. Hence,

$$\eta_k^{AS}(g) = \eta_k^{AS}\left(P^{AS}g\right) = \frac{\frac{1}{(k-1)!}\left|\mu_k\left(P^{AS}g\right)\right|}{\sum_{n=1}^{\infty} \frac{(-1)^n}{(n-1)!} \mu_n\left(P^{AS}g\right)}. \tag{17}$$

This normalization allows us to compare the effect of different moments on the performance measure by taking the ratios of the absolute elasticities for different levels of k. Specifically,

$$\begin{aligned} \frac{\eta_{k+1}^{AS}(g)}{\eta_k^{AS}(g)} &= \frac{\frac{1}{k!}\left|\mu_{k+1}\left(P^{AS}g\right)\right|}{\frac{1}{(k-1)!}\left|\mu_k\left(P^{AS}g\right)\right|} \\ &= \frac{1}{k}\left|\hat{\mu}_{k+1}\left(P^{AS}g\right)\right| \left(\frac{\left|\hat{\mu}_{k+1}\left(P^{AS}g\right)\right|}{\left|\hat{\mu}_k\left(P^{AS}g\right)\right|}\right)^k. \end{aligned} \tag{18}$$

If this elasticity ratio takes a value greater than 1, then the $(k+1)^{th}$ moment has a greater effect on the performance measure than the kth moment. On the other hand, a ratio less than 1 implies that the P^{AS} measure is less sensitive to the $(k+1)^{th}$ moment as compared to the kth moment.

To understand the forces that drive this elasticity ratio to be higher or lower than 1, note first that it consists of three components: the factor $\frac{1}{k}$, $\hat{\mu}_{k+1}\left(P^{AS}g\right)$, and the ratio of $\hat{\mu}_{k+1}\left(P^{AS}g\right)$ to $\hat{\mu}_k\left(P^{AS}g\right)$ raised to the kth power. First, for all $k \geq 1$, we have that $\frac{1}{k} \leq 1$, which drives down the elasticity ratio, and thus the importance of higher moments.

The second component, $\left|\hat{\mu}_{k+1}\left(P^{AS}g\right)\right|$ can be higher or lower than 1 and so can either strengthen or weaken the importance of higher moments. Third and perhaps most interesting is that the third factor introduces a force always making higher moments more important. To see this, note that by Hölder's inequality,

$$\hat{\mu}_{k+1}\left(P^{AS}\left|g\right|\right) \geq \hat{\mu}_{k}\left(P^{AS}\left|g\right|\right). \tag{19}$$

This implies that when k is an odd number,

$$\left|\frac{\hat{\mu}_{k+1}\left(P^{AS}g\right)}{\hat{\mu}_{k}\left(P^{AS}g\right)}\right|^{k} \geq 1. \tag{20}$$

And, when k is even,

$$\left(\frac{\hat{\mu}_{k+2}\left(P^{AS}g\right)}{\hat{\mu}_{k}\left(P^{AS}g\right)}\right)^{k} \geq 1. \tag{21}$$

Thus, the third factor is necessarily greater than 1 for odd values of k. And, for even values of k, we still have a trend up when comparing the $(k+2)^{th}$ to the kth moment.

To summarize, somewhat surprisingly, we do not find that higher moments necessarily have a weaker effect on performance evaluation. Rather, we see forces in either direction. In Section 5 we illustrate this point, showing that higher moments often have a significant effect.

Similarly, for any gamble $g \in \mathcal{G}$ such that $-1 < P^{FH}(g)\, g \leq 1$, we can calculate the absolute elasticity of P^{FH} with respect to $\hat{\mu}_k$. The resulting elasticity ratio is given by

$$\frac{\eta_{k+1}^{FH}(g)}{\eta_{k}^{FH}(g)} = \left|\hat{\mu}_{k+1}\left(P^{FH}g\right)\right| \left(\frac{\left|\hat{\mu}_{k+1}\left(P^{FH}g\right)\right|}{\left|\hat{\mu}_{k}\left(P^{FH}g\right)\right|}\right)^{k}. \tag{22}$$

Note that compared to the case with the P^{AS} measure, the elasticity ratio for the P^{FH} measure only has two components, which correspond to the second and third factors in the P^{AS} case. In particular, now we lose the first factor $\frac{1}{k}$, which serves as a depreciating component in the P^{AS} case. Therefore, the force driving up the importance of higher moments is even stronger with the P^{FH} index.

The discussion above considers the magnitude of the moment effects with respect to raw moments. Similar results are obtained with respect to central moments. We denote such elasticities by $\zeta_{k}^{AS}(g)$ and $\zeta_{k}^{FH}(g)$. For brevity we omit this analysis.

4.2.1. Example: a shifted lognormal distribution

To illustrate the magnitudes of the moment properties and how they affect performance evaluation, consider the following example related to the P^{FH} measure. Let $x \sim N\left(x_0, \sigma^2\right)$ be a normal variable and consider the shifted lognormal gamble $g = \exp(x) - \exp(x_0)$. It is easily verified that $P^{FH}(g) = \exp(-x_0)$. Hence, $P^{FH}(g)$ depends only on x_0 but not on σ^2. Intuitively, σ^2 affects all distribution moments but does not affect P^{FH} since the effects on the different moments offset each other. Moreover, calculation shows that $\eta_1^{FH}(g) = 1$ and $\zeta_2^{FH}(g) > 1$. Thus, if we would increase *only* the first moment by 1%, performance would improve by 1%. And, if we would increase *only* the second moment by 1%, then performance would decline by more than 1%. Also, calculation shows that if σ^2 is sufficiently high, then both $\zeta_3^{FH}(g)$ and $\zeta_4^{FH}(g)$ become larger than 1, and in fact diverge to infinity as σ^2 diverges. Thus, higher distribution moments in this example can have a strong effect on performance.

4.3. Rare disasters

4.3.1. Effect of rare disasters on performance

In some cases gambles feature very bad events that occur with a very small probability. As discussed in Section 2, small probability events are not likely to affect low moments, but may become dominant when high moments are taken into account. Thus, the measures discussed here are well suited to reflect such events. In fact, the two measures differ in the way they account for rare disasters.

An important property of the P^{FH} measure is that it is extremely sensitive to rare disasters [see the discussion in Section V.B in Foster and Hart (2009)]. To formalize this property in our context, let $g_0 \in \mathcal{G}$ be a gamble and choose $L > 0$ very large. One can think of g_0 as a "business as usual" gamble that involves some gains and losses but no disastrous events, whereas $-L$ is a very big and unusual loss. Then, consider the composite gamble g_α that assigns probability $1-\alpha$ to g_0 and α to $-L$, where α is some small probability. The gamble g_α reflects both "business-as-usual" realizations

and the rare disaster. As α becomes very small, the P^{FH} index becomes completely dominated by the disastrous loss L. Namely, $\lim_{\alpha\to 0} P^{FH}(g_\alpha) = 1/L$. Formally,

Proposition 8. *Let $g_0 \in \mathcal{G}$ be a gamble and $L > 0$ such that $P^{FH}(g_0) > 1/L$. Let $\alpha \in (0,1)$ and let g_α denote a composite gamble that assigns probability $1-\alpha$ to g_0 and α to $-L$. Then, $\lim_{\alpha\to 0} P^{FH}(g_\alpha) = 1/L$.*[4]

This follows intuitively from (4). Indeed, the argument of the log function cannot be negative, and thus regardless of the probability assigned to the disaster $-L$, we have $P^{FH} < \frac{1}{L}$. When the probability of the rare disaster becomes small, this inequality becomes more and more binding, as the effect of the "business as usual" gamble g_0 becomes prominent. Thus, with rare disasters the wealth level of a log investor needed to accept the gamble is roughly equal to the worst-case loss, and the P^{FH} measure is roughly equal to $\frac{1}{L}$.

It is important to note that a corresponding result does not hold for the P^{AS} index. In fact, the continuity property in Aumann and Serrano (2008, p. 819) implies that $\lim_{\alpha\to 0} P^{AS}(g_\alpha) = P^{AS}(g_0)$ whenever $\{g_\alpha\}$ are uniformly bounded. In Section 5 we illustrate that indeed, the P^{FH} index is much more sensitive to isolated bad events than the P^{AS} index.

4.3.2. Modeling rare disasters

Measuring, modeling, and estimating rare disasters in practice is a challenge since, by definition, data on such events are scarce. One approach is to consider a single low and rare outcome as a rare disaster. For example, Chen, Joslin, and Tran (2012) model a rare disaster in consumption by taking a single value that matches the calibration of Barro (2006). An alternative approach is to use a distribution of disaster sizes. For example, Barro and Jin (2011) use a power law distribution to model the left tail of both consumption and gross domestic product (GDP).

[4] It is implicit in the statement of the theorem that the sequence $\{g_\alpha\}$ is convergent. Note also that since $g_0 \in \mathcal{G}$, we have that $E(g_0) > 0$. It follows that for all α in a right neighborhood of 0, $E(g_\alpha) > 0$ and thereby, $g_\alpha \in \mathcal{G}$ and $P^{FH}(g_\alpha)$ is well defined.

These two approaches have implications for performance evaluation using P^{AS} and P^{FH}. On one hand consider a single disaster $-L$, and on the other hand replace $-L$ with a distribution that has mean $-L$. Then, conditional on being in the left tail, the latter case imposes a mean-preserving spread relative to the former. By Rothschild and Stiglitz (1970), a mean-preserving spread implies second-order stochastic dominance. Thereby, we have the following corollary of Propositions 1 and 2.

Corollary 1. *Both P^{AS} and P^{FH} favor a gamble with a single disaster $-L$ over a gamble with a distribution of disasters with mean $-L$.*

Thus, using a single disaster size set at the mean of the disaster distribution results in improved performance. But, if the single disaster size is set to be lower than the mean of the disaster distribution, then the performance ranking between the two cases is no longer clear, and it depends on the choice of parameter values. We provide an example in Section 5.3.

4.4. Scale, leverage, and diversification

In applying the two indices one should use caution when dealing with the scale of the gamble and with leverage. To see this point consider a gamble g, and scale it up to αg with $\alpha > 1$. The homogeneity of the indices implies that $P(\alpha g) = \frac{1}{\alpha} P(g) < P(g)$.[5] This is a simple reflection of the fact that if g is rejected by a risk-averse individual compared to the status quo, then αg must also be rejected. Indeed, fix any increasing and concave utility u. Then, by Jensen's inequality for every $\alpha > 1$,

$$\mathbf{E}[u(w_0 + g)] - u(w_0) \geq \frac{1}{\alpha - 1} \left(\mathbf{E}[u(w_0 + \alpha g)] - \mathbf{E}[u(w_0 + g)] \right). \tag{23}$$

Hence, if g is rejected (i.e., $\mathbf{E}[u(w_0 + g)] \leq u(w_0)$), then also αg is rejected (since $\mathbf{E}[u(w_0 + \alpha g)] \leq \mathbf{E}[u(w_0 + g)] \leq u(w_0)$). In words, a scaled up version of a rejected gamble cannot be accepted by a risk-averse investor.

Similarly, if there exists a risk-free asset with return $r_f > 0$, we can consider the

[5] Any time we use P it means that the statement applies to both P^{AS} and P^{FH}.

gamble $\alpha g + (1-\alpha) r_f$, assuming it is in $\mathcal{G}$. Then for $\alpha > 1$,

$$P(\alpha g + (1-\alpha) r_f) \leq P(g). \tag{24}$$

This again reflects that any time a risk-averse individual rejects a gamble g (compared to the status quo), she also rejects a levered version of that gamble. Indeed, we already know that $P(\alpha g) < P(g)$. Then, (24) follows immediately from the fact that $\alpha g + (1-\alpha) r_f$ is first-order stochastically dominated by αg. Given this, for practical applications of the indices it may make sense to compare gambles that have the same scale and leverage. We follow this approach in Section 5.

Finally, we have so far restricted attention to choosing one of two gambles. In practice, there could be multiple gambles available for investment, and an investor may be allowed to invest in a portfolio of gambles. The following proposition establishes that according to the P^{AS} and P^{FH} measures, diversification always (weakly) improves value and that a unique optimal portfolio of any two gambles exists. This follows from convexity properties of the original AS and FH risk measures.

Proposition 9. *Let $g_1, g_2 \in \mathcal{G}$ be two different bounded gambles such that $\alpha g_1 + (1-\alpha) g_2 \in \mathcal{G}$ for all $\alpha \in [0,1]$. Then, there exists a unique $\hat{\alpha}^{AS}$ ($\hat{\alpha}^{FH}$) $\in [0,1]$ that maximizes P^{AS} (P^{FH}) over all convex combinations of g_1 and g_2.*

4.5. Background risk

When investors make decisions, they often face exogenous and unavoidable risks. For example, an investor choosing among different mutual funds may face uncertainty regarding inflation or her labor income. This kind of uncertainty is typically termed "background risk." It is important to know whether our results are sensitive to the presence of such risks. This section explores this issue.

We consider two approaches to modeling background risk. First is "additive background risk" following Eeckhoudt, Gollier, and Schlesinger (1996), who note that exposure to background risk can be modeled as adding a random shock to the investor's initial wealth. Another approach consists of "multiplicative background

risk," where the final wealth of the investor is multiplied by a random shock. The additive approach may fit a shock to the investor's endowment such as due to labor income, whereas the multiplicative approach may fit a random shock to the entire macroeconomic environment, e.g., due to inflation. The P^{AS} measure lends itself naturally to the additive framework, while the P^{FH} measure fits naturally into the multiplicative approach.[6] We discuss each of them separately.

4.5.1. Background risk and the AS measure

Assume that the investor has utility function $u(\cdot)$ and initial wealth w_0. In addition, assume that the investor faces background risk captured by a random variable y, which is independent from any gamble to be considered. The investor needs to determine whether to accept gamble g or not. If she chooses not to take the gamble, then her expected utility is $\mathbf{E}[u(w_0 + y)]$. Alternatively, if she accepts the gamble, her expected utility becomes $\mathbf{E}[u(w_0 + y + g)]$.

For any given utility u and background risk y, we define a new utility function v by $v(w; u, y) = \mathbf{E}[u(w + y)]$. Note that if $u(\cdot)$ is increasing and concave then so is $v(\cdot; u, y)$. Thus, for a given background risk y we can consider the notion of wealth-uniform rejection with respect to utilities of the form $v(\cdot; u, y)$. A difficulty is that in general, $u \in \mathcal{U}^*$ does not imply $v \in \mathcal{U}^*$.[7] Nevertheless, the results related to the P^{AS} measure only require weak DARA, with CARA being a special case [see the proof of Theorem 1 in Hart (2011)]. Let $\mathcal{U}^D$ denote the set of all utility functions satisfying weak DARA. We can show that if $u \in \mathcal{U}^D$, then also $v(\cdot; u, y) \in \mathcal{U}^D$. This enables us to establish the following result.

[6] The reason for this distinction is related to the properties of the exponential and log utility functions. An additive shock can be separated from the wealth of an investor with CARA utility, which serves as a tool for the P^{AS} measure calculation. Similarly, a multiplicative shock can be separated from the wealth of an investor with log utility, which serves as a tool for the P^{FH} measure calculation. These separation properties facilitate the generalization of the measures to allow for background risk.

[7] For example, note that $\log(w + k)$ does not satisfy IRRA for k negative. Thus, if we set $u(w) = \log(w)$ and consider a sequence of background noises $\{y_n\}$ that converges in distribution to the constant $k < 0$, then $v_n(w) = \mathbf{E}[u(w + y_n)]$ fails IRRA for n large enough.

Proposition 10. *Consider an additive background risk y. Gamble g wealth-uniformly dominates gamble g' if and only if $P^{AS}(g) \geq P^{AS}(g')$, where the wealth-uniform dominance order is defined over utility functions of the form $v(\cdot; u, y)$ where $u \in \mathcal{U}^D$.*

This proposition tells us that if investors have weak DARA utility, then additive background risk does not affect their investment rankings when it comes to the P^{AS} measure. Thus, additive background risk can be ignored when using the P^{AS} measure.

4.5.2. Background risk and the FH measure

A multiplicative model of background risk can be obtained by multiplying the final wealth of the investor with a random shock $1 + z \geq 0$. Thus, for any given utility u and background risk $1+z$, we define a new utility function v by $v(w; u, z) = \mathbf{E}[u(w(1+z))]$. Note that if $u(\cdot)$ is increasing and concave, then so is $v(\cdot; u, z)$. We can then consider the notion of utility-uniform rejection applied to utility functions of the form $v(\cdot; u, z)$. As before, $u \in \mathcal{U}^*$ does not necessarily imply $v \in \mathcal{U}^*$. However, to establish our results related to the P^{FH} measure, it is sufficient to consider utility functions with relative risk aversion weakly larger than 1, which naturally includes log utility [see the proof of Theorem 1 in Hart (2011), and in particular, Remark (1) on p. 637]. Let $\mathcal{U}^I$ denote the set of all utility functions with relative risk aversion being at least 1. We can show that if $u \in \mathcal{U}^I$, then also $v(\cdot; u, z) \in \mathcal{U}^I$. This facilitates the following result.

Proposition 11. *Consider a multiplicative background risk $1 + z$. Gamble g utility-uniformly dominates gamble g' if and only if $P^{FH}(g) \geq P^{FH}(g')$, where the utility-uniform dominance order is defined over utility functions of the form $v(\cdot; u, z)$ where $u \in \mathcal{U}^I$.*

Thus, if investors have relative risk aversion weakly larger than 1, multiplicative background risk does not affect their investment rankings using the P^{FH} measure, and hence could be ignored.

5. Applications

Having established the properties of the indices, we now study several applications. We first discuss how the indices can be estimated empirically, and then provide several settings in which the indices are used to draw conclusions about the attractiveness of different investment strategies. The high-moment and rare disaster properties play an important role in these analyses.

5.1. Estimation method

5.1.1. Using return data

In most applications the theoretical notion of a "gamble" will be represented by a finite sample of T return observations, reflecting the percentage change in the investment of the agent. To be consistent we assume that all investments are of \$100, which allows us to treat the percentage rates of return as the actual wealth changes.

Additionally, note that the performance measures do not explicitly account for the opportunity cost of investing in a particular gamble. Thus, in each case we need to specify what we view is the right opportunity cost. For example, in the case of investing in the security market one can view the opportunity cost as investing in a risk-free asset. In that case, the gamble that should be evaluated is that of borrowing \$100 at the risk-free rate and investing this amount in the market. Thus, the returns to be evaluated are, in fact, the excess returns (returns less risk-free rate). Different cases we analyze below will be associated with different natural opportunity costs.

5.1.2. Statistical estimation of the measures

Consider a finite sample of T return observations. As usual in the performance evaluation literature, we assign each observation a probability of $\frac{1}{T}$. Then, we estimate P^{AS} and P^{FH} by solving the implicit equations (1) and (3). Typically, we view the finite sample as a random sample from a population of returns. Then, P^{AS} and P^{FH} which we calculate are just the sample estimates of the "true" performance indices.

To answer questions regarding the underlying population, we apply the Generalized Method of Moments (GMM) (see Hansen, 1982). GMM works out a distribution for the parameter estimates, and, in particular, generates standard errors that we can use to test hypotheses. This approach is natural since (1) and (3) can directly be viewed as moment conditions, and thus lend themselves easily to GMM estimation.[8] Notice that in either case, we have exactly one parameter to be estimated and one moment condition, implying that our problem is "just identified," and the GMM estimate is determined such that the sample average of the moment condition equals zero. Thus, the GMM estimates equal the solutions to the implicit equations (1) and (3). These estimates are consistent and asymptotically normal. Moreover, since the model is just identified, the weighting matrix used for the GMM estimation is irrelevant, and the estimates are efficient. We also obtain a covariance matrix, which yields standard errors that allow us to examine the statistical significance of the measures.

Often we compare the performance indices for multiple gambles to assess which investment strategy dominates. We then estimate the performance indices jointly. That is, we estimate the performance indices of n gambles using the n moment conditions implied by the implicit equations (1) (or (3)). The resulting covariance matrix allows us to compute a standard error for the difference in the measures of any two gambles, which can then be used to determine whether the gambles generate significantly different performance indices in the population.

To be consistent with the estimation of the performance indices, we also use the GMM approach to estimate various moments of gambles. It is worth noting that the GMM estimates of central moments are biased, but this has a negligible effect on our results due to the consistency of the estimates.

[8] Bali, Cakici, and Chabi-Yo (2011) offer an alternative way to estimate the measures using option prices. Such an approach may be applied for assets with available liquid options. Note also that their calculations of the risk measures apply risk-neutral probabilities, although the measures are stated in terms of physical probabilities. Thus, their calculations tend to overweight the probabilities of bad outcomes, essentially correcting for risk twice. Advantages of the GMM approach proposed here is its wide applicability, the fact that it allows one to test hypotheses, and its direct reliance on physical probabilities.

5.2. Application I: attractiveness of anomalies

As a first application of the indices, we evaluate the attractiveness of popular investment strategies that rely on well-documented anomalies. It is well established that small firms (those with low market capitalizations), and value firms (those with high book-to-market ratios) gain abnormal average returns in US equity markets (see Banz, 1981; Rosenberg, Reid, and Lanstein, 1985; Fama and French, 1992). Additionally, it is established that momentum strategies, i.e., holding long positions in stocks that yielded high returns in the recent past while holding a short position in stocks that yielded low returns in the recent past, generate abnormal returns (see Jegadeesh and Titman, 1993). Are these trading strategies still attractive when accounting for their high-moment properties?

To evaluate this issue we use the Fama and French (1993) and Carhart (1997) portfolios, which are constructed based on these anomalies. Specifically, Fama and French (1993) and Carhart (1997) construct four portfolios. The first is denoted $mktrf$ (for market less the risk-free rate). The returns for this portfolio reflect an investment in a well-diversified portfolio of US stocks, where the opportunity cost is assumed to be investment in a risk-free asset. The second portfolio is smb (for small minus big). This portfolio is long in low market-capitalization stocks and short in high market-capitalization stocks. Historically, this portfolio yielded abnormal returns reflecting what is known as the "small firm anomaly." Note that the underlying assumption here is that the opportunity cost of investing in small stocks is the return of investing in large stocks. The third portfolio is hml (for high minus low). This portfolio is long in high book-to-market stocks (value stocks) and short in low book-to-market stocks (growth stocks). This portfolio builds on the "value anomaly." The fourth portfolio is umd (for up minus down). This portfolio is long in stocks that had high returns in the recent past (winners) and short in stocks that had low returns in the recent past (losers), building on the "momentum anomaly."

We obtain monthly data for the four portfolios for the period January 1962 to December 2009 from Kenneth French's data library. Panel A of Table 3 reports

summary statistics for the first moment (μ_1) and the three higher central moments (m_2, m_3, and m_4) for the four portfolios. The table also compares the moments for the different portfolios using GMM standard errors.

The average monthly market excess return ($mktrf$) during our sample period is 0.41%, the average monthly return on the *smb* portfolio is 0.23%, and the average monthly returns on the *hml* and *umd* portfolios are 0.44% and 0.73%, respectively. These averages are consistent with prior studies, and reflect the popular attractiveness of the value (*hml*) and momentum (*umd*) anomalies. In terms of higher moments we see that $mktrf$ has the highest variance (m_2), while *umd* is the most negatively skewed and exhibits the largest tail-risk (m_4) out of the four portfolios. Fig. 1 presents histograms of the returns of the four portfolios. One interesting feature that can be learned from this figure is that the *umd* portfolio has some extreme and rare bad events. These "rare disasters" contribute to the high tail-risk associated with this strategy and perhaps also to the left skewness.

Panel B of Table 3 reports the Sharpe ratios and the P^{AS} and P^{FH} performance indices.[9] The Sharpe ratio of the *umd* portfolio appears to be the highest among the four, with the Sharpe ratio of the *smb* the lowest, although the differences are not statistically significant. Both the P^{AS} and P^{FH} measures, however, suggest that *hml* is the superior portfolio. Apparently, the high negative skewness and tail-risk of *umd* lower its attractiveness. Indeed, comparing the *umd* and *hml* portfolios, we see that the former has more negative skewness (-116.1 vs. -0.67) and higher tail-risk (4,862.9 vs. 390.8). Thus, despite the fact that the Sharpe ratio of *umd* is slightly higher than that of *hml*, the P^{AS} and P^{FH} measures favor *hml*.

It is interesting to note that the negative rare events showing in the distribution of *umd* play an important role in the determination of P^{FH}. Recall from Proposition 8 that in the presence of a rare disaster, we have $P^{FH} \approx \frac{1}{L}$, where L is the loss in the case of the disaster. In the case of *umd*, Fig. 1 shows a "disaster return" of

[9] We estimate Sharpe ratios and their associated standard errors following Lo (2002). This is achieved by applying the delta method based on the GMM estimates of the mean and the variance of the portfolio returns.

-34.7%. This reflects the return on momentum investing in April 2009, during which the market experienced a sharp reversal. Accordingly, we have that P^{FH} for the *umd* portfolio is $\frac{1}{34.7} \approx 0.029$. Thus, in the case of *umd*, the rare disaster dominates the performance measurement of the portfolio, when using P^{FH}.

Panel B of Table 3 also reports the elasticities, which reflect the importance of the different moments in the determination of the performance measures (as discussed in Section 4.2). Recall, in particular, that the importance of the moments may or may not be monotone. Thus, higher moments can potentially be very influential in determining performance. The elasticities reported in the table suggest that the first two moments have a strong effect on performance of the four portfolios. The elasticities of the third and fourth moments are lower but still meaningful. And, as suggested in Section 4.2, the elasticities are often non-monotone. For example, for the P^{AS} measure of the *hml* portfolio, the fourth moment has a roughly 20 times larger elasticity compared to the third moment, due to the very small skewness of this portfolio. Note that in the case of the P^{FH} measure of the *umd* portfolio, all reported elasticities are very small. This is a reflection of the fact that in this case the P^{FH} measure is dominantly determined by the rare disaster.

Panel B also reports the certainty equivalent of a risk-averse investor with CRRA utility given by $u(w) = \frac{w^{1-\gamma}}{1-\gamma}$. For illustration we use $\gamma = 3, 5, 10$, which are levels of risk aversion commonly used in asset pricing calibrations. In addition, we set the initial wealth and the scale of investment to be equal. For example, the results suggest that when $\gamma = 5$, investments in the *mktrf* and *smb* portfolios are equivalent to certain -0.14% and -0.02% monthly losses, while investments in the *hml* and *umd* portfolios are equivalent to certain 0.23% and 0.14% monthly gains, respectively. However, the differences in the certainty equivalent for the four portfolios are not statistically significant in most cases. Also notice that the *hml* portfolio becomes more attractive relative to the *umd* portfolio as the level of risk aversion increases.

In Section 4.4 we have established that diversification improves performance according to both P^{AS} and P^{FH}, and that for any two gambles there exists a unique

convex combination that maximizes each of the performance measures (Proposition 9). To illustrate this result we now study how combinations of the four Fama-French and Carhart portfolios affect their attractiveness. Specifically, for each pair of the four portfolios, we numerically find the optimal weight maximizing P^{AS} and P^{FH}. The results are reported in Table 4, in which for each pair of portfolios we report the maximal performance and the associated optimal weights (in parentheses).

As expected, the performance measures for the optimal convex combinations are always higher than those for the associated "stand alone" portfolios. For example, Panel B of Table 3 shows that the P^{AS} measures for the $mktrf$ and smb portfolios are 0.0382 and 0.0468, respectively. In comparison, the optimal convex combination consisting of 35% invested in $mktrf$ and 65% invested in smb has a P^{AS} value of 0.0659, which is more than 40% higher than the individual performances. The best pairwise performance according to both measures is achieved by mixing between smb and hml. Specifically, the optimal combination for the P^{AS} measure is obtained by investing 43% in smb and 57% in hml, which delivers a P^{AS} value of 0.2006. In comparison, if one mixes between hml and umd, then the maximal P^{AS} value is 0.1933. Similarly, the optimal combination for the P^{FH} measure consists of 54% and 46% invested in smb and hml, respectively. The associated maximal P^{FH} value is equal to 0.1540, which is higher than 0.1143, the maximal P^{FH} value achieved by mixing between hml and umd.

In summary, the discussion in this section illustrates that high moments and rare disasters may have a meaningful effect on the performance evaluation of popular trading strategies. In particular, the popular momentum strategy appears much less attractive using the new performance indices. Indeed, this strategy has high negative skewness, high tail-risk, and it exhibits rare disasters, all of which tend to lower its P^{AS} and P^{FH} performance indices. In addition, the momentum strategy remains dominated even after allowing for diversification between different anomalies.

5.3. Application II: more on the importance of high moments and rare disasters

In this section we demonstrate further the importance of high moments in evaluating the performance of investment portfolios. To illustrate this point we first use the momentum (*umd*) portfolio presented above. We ask the following question: To what extent would performance evaluation be biased if we accounted for the first two moments only, essentially ignoring the effects of higher moments? To answer this question we draw one million samples of 576 realizations (same size as the *umd* sample) from a simulated normal distribution with mean and variance equal to those of the sample momentum returns. We also draw one million samples with replacement from the original *umd* portfolio return distribution. Fig. 2 shows the distributions of the *umd* portfolio and one arbitrary sample simulated using the normal distribution. It appears that the simulated version of *umd* is less left skewed and does not suffer from rare disasters. To formally compare the two distributions, we calculate bootstrapped standard errors for the two sets of samples.[10]

Panel A of Table 5 reports the moments of the true and simulated *umd* distributions. Observe that the first two moments are very close (as expected). The higher moments look quite different. For example, the fourth moment of *umd* is 4,863 compared to 1,055 for the simulated distribution. However, the bootstrapped standard errors are quite large, and we cannot reject the null that the two are statistically equal. Panel B of Table 5 reports the Sharpe ratio and the P^{AS} and P^{FH} performance indices. The Sharpe ratio is close for the true and simulated portfolios, which is expected as both portfolios have similar mean and variance. The P^{AS} index is lower for the *umd* portfolio as compared to the simulated one (0.063 vs. 0.078), although the difference is not statistically significant. The P^{FH} index significantly favors the simulated normal distribution, as it is about 2.3 times higher than that for the *umd* portfolio (significant at the 1% level). The difference in the P^{FH} indices

[10] One could alternatively use our GMM standard errors in this case. None of the conclusions below depends on this choice.

comes largely from the existence of a "rare disaster" for the *umd* portfolio. In sum, this example shows that relying on the first two moments only may lead one astray.

We next illustrate how different approaches to modeling rare disasters can affect performance evaluation. Section 4.3.2 discussed two such modeling approaches: (i) using a single disaster size; or (ii) applying a distribution of rare disasters. To construct a parametric example, we apply these two approaches to the Fama-French and Carhart factor portfolios.

For each portfolio we define monthly return realizations falling below the 1% quantile as rare disasters. In other words, these "disastrous" events occur about once every 8.3 years. The values of these disaster thresholds in our sample are -11.8%, -6.8%, -8.5%, and -11.6% for $mktrf$, smb, hml, and umd, respectively. We then calculate for each portfolio the average disaster size conditional on the occurrence of a disastrous event. These averages in our sample are -15.9%, -9.5%, -9.9%, and -19.0% for $mktrf$, smb, hml, and umd, respectively. Next, we compare the P^{AS} and P^{FH} measures for the following two scenarios. In one scenario, the measures are computed using the original sample returns, which is exactly what we did before. In another, we replace all disastrous realizations by the single average disaster size. As expected from Corollary 1, both measures are higher in the case of single disaster size. The difference is statistically insignificant in most cases, but it is both economically and statistically significant at the 1% level for the P^{FH} measure of the *umd* portfolio. This reflects the sensitivity of P^{FH} to rare disasters and the existence of a large loss for the *umd* portfolio.

5.4. Application III: private vs. public equity

In our next illustration of the performance indices, we compare the performance of private and public equity. Moskowitz and Vissing-Jorgensen (2002) provide a thorough comparison of the performance of public vs. private equity investment from the point of view of individual investors. They find that the returns to private equity are not higher than those of public equity, when accounting for survival. This

result is puzzling since private equity investments expose investors to a high level of idiosyncratic risk. They observe that private equity investment is right skewed and conjecture that preference for skewness may be one reason for the tendency of individuals to invest in private equity.

The performance indices used in this paper account for all moments of the distribution (skewness in particular). We thus use the indices to explore whether the right skewness of private equity indeed compensates for its otherwise unattractive performance. Our method follows Moskowitz and Vissing-Jorgensen (2002). Using data on individual household investment in private equity from the 2004 Survey of Consumer Finance (SCF), we estimate excess returns obtained by households since the founding or the acquisition of a private firm. These are the returns they would achieve by borrowing at the risk-free rate and investing in the private firm. We consider only private firms in which a household has its largest actively managed equity position. We treat each household as an observation and estimate the average annual holding period return. We disregard observations with a holding period of less than one year. The average annual holding return is calculated as the sum of the geometric average annual capital gain and the current dividend return, which is assumed to be representative of those in previous years. While the former can be estimated from the initial and current market value of the ownership share, the latter is computed from the earnings in the year prior to the survey and the current market value of equity assuming that 30% of earnings are retained in the firm. Finally, we subtract the risk-free rate from the estimated holding period return to obtain the average annual excess return. Notice that similar to Moskowitz and Vissing-Jorgensen (2002), our sample is conditional on survival of the private firms, yielding an upward bias in the evaluation of private equity performance.

For the purpose of comparison, we compute for each household the geometric average annual return it would obtain by investing in the Center for Research in Security Prices (CRSP) value-weighted market index for the same time period as its private equity holdings. As before, we subtract the risk-free rate to get the excess

return. Fig. 3 shows the distributions of private and public equity. For the sake of plotting the private equity histogram, we winsorize the excess returns at 200%. Importantly, we do not winsorize the data in any of our analyses, since extreme events are key to our results. In line with the results of Moskowitz and Vissing-Jorgensen (2002), the figures suggest that private equity returns are much more dispersed and right skewed compared to public equity.

Panel A of Table 6 reports the moments for the private and public equity returns. The average return on private equity conditional on survival is about 35 times larger than that of public equity, and the difference is significant at the 1% level. However, private equity is also extremely more volatile. In addition, the third moment of private equity is much larger than that of public equity, but it also exhibits much heavier tails as reflected in the fourth moment. Overall, based on these four moments, it is not clear a priori which of the two dominates.

Panel B of Table 6 provides the calculations of the performance indices. The Sharpe ratio of private equity is significantly lower than that of public equity. That is, the superior average returns of private equity are outweighed by their high volatility. The question is whether the high-moment properties of private equity make it attractive [as conjectured by Moskowitz and Vissing-Jorgensen (2002)]. The answer provided by P^{AS} and P^{FH} is negative. Both measures are significantly higher for public equity as opposed to private equity, even when conditioning on survival. The right skewness of private equity is not sufficient to compensate for its other moments. In other words, the "private equity premium puzzle" suggested by Moskowitz and Vissing-Jorgensen (2002) does not seem to be resolved by high-moment properties.

5.5. Application IV: active vs. passive mutual funds

In the next illustration of the indices we compare the performance of actively managed equity funds and index equity funds. There is a long-lasting debate on the value of active as opposed to passive management of mutual funds (e.g., Jensen, 1968; Henriksson, 1984; Chang and Lewellen, 1984; Ippolito, 1989; Malkiel, 1995; Wermers,

2000). The focus of this research is on the first and second moments, i.e., on whether active management beats the average returns of passive management accounting for risk, as traditionally perceived. Given that investors care about all moments of the return distribution, we extend the analysis to account for those moments using the new performance measures.

To examine this issue we obtain mutual fund return data from CRSP. This data set includes the past records of all open-ended mutual funds in existence, and is thus free of survivorship bias. The returns we use are net of fees. We also obtain mutual fund identification data from the Mutual Fund Links (MFLinks) data set maintained by the Wharton Research Data Services (WRDS). Our sample period for this analysis is January 1991 to December 2009.[11]

The MFLinks data provide a reliable means to join the CRSP Mutual Fund data to equity holdings information. In practice, an investment company may run different portfolios that have the same composition but serve different groups of investors. Such portfolios are treated as separate funds in CRSP as noted in Wermers (2000). The MFLinks allows us to combine all these portfolios and treat them as one mutual fund.

To evaluate the performance of actively managed funds, we drop all index funds (including index-based and enhanced index funds), bond funds, and balanced funds and keep only the actively managed equity funds. Also excluded from our analysis are funds investing in foreign securities and sector funds that invest in particular industries, such as utilities and real estate. Similarly, we obtain a sample of index equity funds by dropping all actively managed funds, bond funds, balanced funds, foreign funds, and sector funds.[12] We then compute for each month during the sample period the value-weighted average return of all actively managed funds and index funds, and we subtract the risk-free rate to obtain excess returns (228 observations overall).

[11] We begin our sample in 1991 since before that time, we do not have accurate information distinguishing between index funds and actively managed funds.

[12] In our sample of index funds, we do not include index-based and enhanced index funds. Including those in the analysis does not change the results materially and the conclusions are unaffected.

Panel A of Table 7 reports summary statistics for both groups. It can be seen that all four moments are very similar for the active and passive funds, and the differences between them are not statistically significant. Panel B reports the Sharpe ratio and the P^{AS} and P^{FH} indices. Here again, we do not see significant differences across the two fund groups, consistent with the similar moments reported in Panel A. Thus, the P^{AS} and P^{FH} indices lead to the conclusion that after-fee returns of active and index funds are not distinguishable, even after accounting for high moments. This result reinforces the view that active management does not improve overall performance of mutual funds. In fact, we find that not only is overall performance not affected by active management, but also each of the first four return distribution moments are not affected by active management.

6. Mutual fund selection based on the performance indices

So far we have illustrated how the new indices can be used to evaluate the performance of investments. In this section we take one step forward and examine the performance of investments selected based on the new indices out of sample. The idea is that if the high-moment properties of investment portfolios are persistent, then portfolios sorted on historical AS and FH measures would exhibit superior performance in the future. To this end, we use the new performance indices to select mutual funds on a monthly basis, and then examine the outcome of such an investment strategy.

Our data for this exercise consist of monthly returns of all mutual funds from CRSP and identification data from MFLinks. We focus on a sample period covering 576 months from January 1962 through December 2009. The original CRSP data set covers 44,093 funds. We drop all index funds, bond funds, and balanced funds and keep only the actively managed equity funds. We also exclude foreign funds and sector funds. This leaves us with 15,580 actively managed mutual funds that invest in domestic equity securities. After combining duplicate funds using the MFLinks

data we are left with 3,222 unique funds.

Our aim is to examine the performance of mutual-fund portfolios selected based on the new indices as compared to those selected based on a traditional measure (the Sharpe ratio). In each month during our sample period starting from January 1967 and for each mutual fund, we calculate the P^{AS}, P^{FH} indices and the Sharpe ratio (S) based on the most recent 60 monthly excess returns assigning equal probability (1/60) to each observation. If a fund does not have complete records for the preceding 60 months, or if the average return is not positive during that period, we disregard it for that particular month.[13] We then rank all mutual funds in each month based on their indices (separately for P^{AS}, P^{FH}, and S). This yields three portfolios of "selected" mutual funds consisting of equal-weighted combinations of the top decile mutual funds. We rebalance these portfolios on a monthly basis.

Panel A of Table 8 compares various moments for the portfolio excess returns obtained by adopting the three portfolios described above (denoted for brevity by AS, FH, and S), as well as for the market (denoted by MKT, and represented by the CRSP value-weighted index). For each of the four portfolios the table reports return distribution moments estimated using GMM.

The average return (μ_1) appears highest for portfolio S (0.54%), which was constructed based on Sharpe ratios. Indeed, the average returns on the AS and FH portfolios are estimated at 0.50% and 0.47%, respectively, and the average return on the market is estimated at 0.43%. However, the differences are insignificant in most cases. By contrast, higher moments generated by the AS and FH indices are significantly more appealing than those generated by the Sharpe ratio and are also often more appealing than those generated by the market. Consider first the second central moment (m_2). The estimated values for the AS and FH portfolios are 16.2 and 15.7, respectively. Those are significantly lower than the estimates for the S and MKT portfolios, which are 21.2 and 21.4, respectively. Now, consider the third central

[13] Recall that the P^{AS} and P^{FH} measures as well as the Sharpe ratio can be calculated only if the average return is positive. In some cases following very bad years, this prevents us from calculating the measures for some funds over a 60-month period.

moment (m_3). Its estimates for the AS and FH portfolios are -37.3 and -37.0, respectively, as opposed to -59.1 and -55.6 for the S and market portfolios, respectively, with the difference being significant when compared to the S portfolio. Thus, the AS and FH indices appear to generate less negatively skewed returns compared to the S portfolio. Finally, both the AS and the FH portfolios generate a significantly lower fourth central moment (m_4) compared to both the S and MKT portfolios. Thus, the AS and FH portfolios seem to present investors with a lower tail-risk when compared to the market or the Sharpe ratio-based portfolios.

The two performance indices estimated for the four different portfolios (reported in Panel B) encapsulate all the information contained in high moments. Indeed, both P^{AS} and P^{FH} are higher for the AS and FH portfolios compared to S and MKT portfolios, where the difference is statistically significant in most cases. Interestingly, portfolio S, which is composed based on the Sharpe ratio, has a lower Sharpe ratio compared to the AS and FH portfolios, though the differences are not statistically significant.

As before, we also calculate the certainty equivalent obtained from investing in each one of these portfolios through the eyes of a CRRA investor with several levels of risk aversion. Suppose, for instance, that the relative risk aversion coefficient is equal to 5. Then, the results suggest that borrowing at the risk-free rate and investing in the AS and FH portfolios are equivalent to certain 0.07% and 0.05% monthly gains, respectively. In comparison, borrowing at the risk-free rate and investing in the S and MKT portfolios are equivalent to certain -0.03% and -0.15% monthly losses, respectively. Generally, the certainty equivalent generated by the AS and FH portfolios is higher than that generated by the S portfolio and the market. The differences become more significant as the risk aversion coefficient becomes larger.

7. Conclusion

In this paper we explore two measures of performance, P^{AS} and P^{FH}, based on a reinterpretation of two riskiness measures proposed by Aumann and Serrano (2008)

and Foster and Hart (2009). We extend these indices to multi-period investments and investigate their moment properties. We establish that they reflect all distribution moments in a manner consistent with economic intuition and with the asset pricing literature. Namely, they are increasing in mean and skewness and decreasing in variance and tail-risk of the investment return. We also discuss the way these two indices reflect disaster risk.

We then apply these indices to popular investment strategies and to well-known anomalies. We first find that the momentum strategy, which is often considered the most serious and hard to explain deviation from market efficiency, is not an attractive investment strategy according to our performance measures due to unattractive higher moments. Then we show that even after accounting for higher moments, private equity investments are dominated by public equity. Hence, the "private equity premium puzzle" suggested by Moskowitz and Vissing-Jorgensen (2002) is not resolved by high-moment properties. Furthermore, we apply these measures to the comparison of actively managed equity funds and index funds and find that the new performance indices reinforce the view that active management does not improve investment performance.

Finally, we examine whether the new performance measures are helpful with constructing investment strategies. Specifically, we compare the performance of portfolios of mutual funds selected based on the two indices to the market portfolio and to a portfolio selected based on the Sharpe ratio. We find that the P^{AS} and P^{FH} indices give rise to portfolios with moments more appealing than the market portfolio and the portfolio generated by the Sharpe ratio, and that they also produce lower disaster risk.

The measures introduced by Aumann and Serrano (2008) and Foster and Hart (2009) appear to be very useful for evaluating both risk and performance. Future research will likely further explore the implications of these measures for asset pricing, portfolio selection, and for dynamic decision making under risk. Such an approach is consistent with the recent trend in asset pricing of departing from traditional utility

functions, and accounting for higher distribution moments and rare disasters.

Appendix A

Proof of Propositions 3 and 4. We first prove the following lemma, which generalizes Proposition 2 in Hart (2011) to a multi-period setting.

Lemma 1. *Let U_1 and U_2 be two time separable utility functions with $u_1(w)$ and $u_2(w)$ being the associated utility per period. Let $A_{u_1}(w)$ and $A_{u_2}(w)$ denote the absolute risk aversion coefficients of u_1 and u_2, respectively, and let $I \subset (0, \infty)$ be an interval where $A_{u_1}(w) \geq A_{u_2}(w)$ for every $w \in I$. Then for every $w_0 > 0$ and $g \in \mathcal{G}^T$ such that $w_0 + g^t \subset I$ for all $t = 1, 2, ..., T$, if U_2 rejects g at w_0, then U_1 also rejects g at w_0.*

Proof of Lemma 1. We want to prove that if

$$\sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[u_2\left(w_0 + g^t\right)\right] \leq u_2(w_0) \sum_{t=1}^{T} \rho^{t-1}, \tag{25}$$

then we have

$$\sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[u_1\left(w_0 + g^t\right)\right] \leq u_1(w_0) \sum_{t=1}^{T} \rho^{t-1}.$$

It is shown in the proof of Proposition 2 in Hart (2011) that $A_{u_1}(w) \geq A_{u_2}(w)$ for every $w \in I$ if and only if there exists an increasing and concave function ψ such that

$$u_1(\cdot) = \psi(u_2(\cdot)).$$

Now suppose (25) is true. Then this implies

$$\begin{aligned}
& \sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[u_1\left(w_0+g^t\right)\right] \\
= & \sum_{t=1}^{T} \rho^{t-1} \mathbf{E}\left[\psi\left(u_2\left(w_0+g^t\right)\right)\right] \\
\leq & \sum_{t=1}^{T} \rho^{t-1} \psi\left(\mathbf{E}\left[u_2\left(w_0+g^t\right)\right]\right) \quad \text{(by Jensen's inequality)} \\
= & \sum_{q=1}^{T} \rho^{q-1}\left[\sum_{t=1}^{T} \frac{\rho^{t-1}}{\sum_{q=1}^{T} \rho^{q-1}} \psi\left(\mathbf{E}\left[u_2\left(w_0+g^t\right)\right]\right)\right] \\
\leq & \sum_{q=1}^{T} \rho^{q-1} \psi\left(\sum_{t=1}^{T} \frac{\rho^{t-1}}{\sum_{q=1}^{T} \rho^{q-1}} \mathbf{E}\left[u_2\left(w_0+g^t\right)\right]\right) \quad \text{(by Jensen's inequality)} \\
\leq & \sum_{q=1}^{T} \rho^{q-1} \psi\left(\frac{u_2\left(w_0\right) \sum_{t=1}^{T} \rho^{t-1}}{\sum_{q=1}^{T} \rho^{q-1}}\right) \quad \text{(from (25))} \\
= & \sum_{t=1}^{T} \rho^{t-1} \psi\left(u_2\left(w_0\right)\right) \\
= & u_1\left(w_0\right) \sum_{t=1}^{T} \rho^{t-1},
\end{aligned}$$

as stated. □

The existence and uniqueness of the solution follow from the proofs of Theorem A in Aumann and Serrano (2008) and Theorem 1 in Foster and Hart (2009), respectively. The proofs for the correspondence with wealth-uniform and utility-uniform dominance are similar to the proof of Theorem 1 in Hart (2011) with the reference to Proposition 2 in Hart (2011) replaced by Lemma 1 established above. For brevity we do not repeat the details here. □

Proof of Proposition 5. The proofs for the two measures are parallel. Hence, for brevity we only present the proof for the P^{AS} measure. We begin with the following lemma.

Lemma 2. *For any* $g = \left(g^1, g^2\right) \in \mathcal{G}^2$ *and* $P = P^{AS}$ *or* $P = P^{FH}$, *we have* $\min\left\{P\left(g^1\right), P\left(g^2\right)\right\} \leq P(g) \leq \max\left\{P\left(g^1\right), P\left(g^2\right)\right\}$ *with equalities holding if and*

only if $P\left(g^1\right) = P\left(g^2\right)$.

Proof of Lemma 2. As mentioned above, we only present the proof for the P^{AS} measure. For any $p > 0$ and $g \in \mathcal{G}$, let

$$f^{AS}(p, g) \equiv \mathbf{E}[\exp(-pg)] - 1. \tag{26}$$

Then (8) with $T = 2$ can be rewritten as

$$F^{AS}\left(P^{AS}(g), g^1, g^2\right) \equiv f^{AS}(P^{AS}(g), g^1) + \rho f^{AS}(P^{AS}(g), g^2) = 0. \tag{27}$$

If $P^{AS}\left(g^1\right) = P^{AS}\left(g^2\right)$, then by choosing $P^{AS}(g) = P^{AS}\left(g^1\right) = P^{AS}\left(g^2\right)$ one can easily verify that (27) holds. Thus, by uniqueness we have that $P^{AS}(g) = P^{AS}\left(g^1\right) = P^{AS}\left(g^2\right)$.

Now consider the case in which $P^{AS}\left(g^1\right) \neq P^{AS}\left(g^2\right)$. By the proof of Theorem A in Aumann and Serrano (2008), we know that $f^{AS}(\cdot, g^t)$ single-crosses zero at $P^{AS}\left(g^t\right)$ from below for $t = 1, 2$. That is, $f^{AS}(p, g^t) < 0$ for any $p < P^{AS}\left(g^t\right)$, and $f^{AS}(p, g^t) > 0$ for any $p > P^{AS}\left(g^t\right)$.

If $P^{AS}(g) \leq \min\left\{P^{AS}\left(g^1\right), P^{AS}\left(g^2\right)\right\}$, then $f^{AS}(P^{AS}(g), g^t) \leq 0$ for both $t = 1, 2$ with at least one inequality being strict. This contradicts (27). If $P^{AS}(g) \geq \max\left\{P^{AS}\left(g^1\right), P^{AS}\left(g^2\right)\right\}$, then $f^{AS}(p, g^t) \geq 0$ for both $t = 1, 2$ with at least one inequality being strict, which again contradicts (27). Therefore, we must have

$$\min\left\{P^{AS}\left(g^1\right), P^{AS}\left(g^2\right)\right\} < P^{AS}(g) < \max\left\{P^{AS}\left(g^1\right), P^{AS}\left(g^2\right)\right\}. \square$$

Now, implicitly differentiating (27) yields

$$\frac{\partial P^{AS}(g)}{\partial \rho} = -\frac{\partial F^{AS}/\partial \rho}{\partial F^{AS}/\partial P^{AS}(g)} = -\frac{f^{AS}(P^{AS}(g), g^2)}{\partial F^{AS}/\partial P^{AS}(g)}.$$

Similar to the one-period case, $F^{AS}(\cdot, g^1, g^2)$ crosses zero at $P^{AS}(g)$ from below. Hence, we have

$$\frac{\partial F^{AS}}{\partial P^{AS}(g)} > 0.$$

By Lemma 2, the sign of $f^{AS}(P^{AS}(g), g^2)$ is equal to the sign of $P^{AS}(g^1) - P^{AS}(g^2)$.[14] Hence, the sign of $\frac{\partial P^{AS}(g)}{\partial \rho}$ is equal to the sign of $P^{AS}(g^2) - P^{AS}(g^1)$. □

Proof of Proposition 6. We provide the proof for the first part of the proposition and for the case that k is an odd number. The proof of all the other cases is similar. For any $p > 0$ and $g \in \mathcal{G}$, define

$$f^{AS}(p, g) \equiv \mathbf{E}[\exp(-pg)] - 1 = \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} p^n \mu_n(g).$$

By the proof of Theorem A in Aumann and Serrano (2008), we know that $f^{AS}(\cdot, g)$ single-crosses zero at $p = P^{AS}(g) > 0$ from below.

Now, suppose that $g, g' \in \mathcal{G}$ such that $\mu_k(g) > \mu_k(g')$. Since k is odd, the coefficient of $\mu_k(g)$ in $f^{AS}(p, g)$ is negative. Hence,

$$f^{AS}(p, g) < f^{AS}(p, g') \tag{28}$$

for all $p > 0$. By Eq. (12) we know that

$$f^{AS}(P^{AS}(g), g) = 0. \tag{29}$$

Then, by (28) and (29) we have

$$0 = f^{AS}(P^{AS}(g), g) < f^{AS}(P^{AS}(g), g').$$

Since $f^{AS}(\cdot, g')$ single-crosses zero at $P^{AS}(g')$ from below, it must be that $P^{AS}(g') < P^{AS}(g)$. □

Proof of Proposition 7. The proof is similar to that of Proposition 6. □

Proof of Proposition 8. The composite gamble g_α converges in distribution to g_0 as α goes to zero. Hence, by Proposition 10 in Foster and Hart (2009, p. 810), $\lim_{\alpha \to 0} P^{FH}(g_\alpha) = 1/L$. □

[14] If $P^{AS}(g^1) < P^{AS}(g^2)$, we know from Lemma 2 that $P^{AS}(g) < P^{AS}(g^2)$. Since $f^{AS}(\cdot, g^2)$ crosses zero at $P^{AS}(g^2)$ from below, we have $f^{AS}(P^{AS}(g), g^2) < 0$. The other case is parallel.

Proof of Proposition 9. We first prove the result for the P^{AS} measure. For any given $\alpha \in [0,1]$, let

$$h^{AS}(\alpha) = R^{AS}(\alpha g_1 + (1-\alpha) g_2),$$

where $R^{AS}(\cdot)$ is the Aumann-Serrano riskiness index. According to Aumann and Serrano (2008), $R^{AS}(\cdot)$ is subadditive and homogeneous of degree 1, which together implies that $R^{AS}(\cdot)$ is a convex risk measure. That is, $\forall \alpha \in [0,1]$ we have

$$R^{AS}(\alpha g_1 + (1-\alpha) g_2) \leq \alpha R^{AS}(g_1) + (1-\alpha) R^{AS}(g_2).$$

This in turn implies that $h^{AS}(\cdot)$ is a convex function.

On the other hand, according to the continuity property of R^{AS} [see Section V.D of Aumann and Serrano (2008)], when $\alpha_n \to \alpha$, $R^{AS}(\alpha_n g_1 + (1-\alpha_n) g_2) \to R^{AS}(\alpha g_1 + (1-\alpha) g_2)$. This follows because $\alpha_n g_1 + (1-\alpha_n) g_2$ is a uniformly bounded sequence of gambles converging in distribution to $\alpha g_1 + (1-\alpha) g_2$. This implies that $h^{AS}(\cdot)$ is continuous on $[0,1]$. Since a continuous convex function on a bounded domain must have a unique global minimum, there exists a unique $\hat{\alpha}^{AS} \in [0,1]$ that minimizes $h^{AS}(\alpha)$. Since $R^{AS}(\cdot)$ is positive, and

$$P^{AS}(\alpha g_1 + (1-\alpha) g_2) = \frac{1}{R^{AS}(\alpha g_1 + (1-\alpha) g_2)} = \frac{1}{h^{AS}(\alpha)},$$

we have that the same $\hat{\alpha}^{AS}$ maximizes $P^{AS}(\alpha g_1 + (1-\alpha) g_2)$.

The proof for the P^{FH} measure is similar with the only difference being in the continuity property. According to Proposition 3 in Foster and Hart (2009), the continuity of $R^{FH}(\cdot)$ requires not only convergence in distribution but also convergence of the maximal loss for uniformly bounded sequences of gambles. It is clear when $\alpha_n \to \alpha$, the maximal loss of $\alpha_n g_1 + (1-\alpha_n) g_2$ converges to that of $\alpha g_1 + (1-\alpha) g_2$, which establishes the result. □

Proof of Proposition 10. Since y is independent of g, it is immediate that when $u(\cdot)$ exhibits CARA, $v(\cdot; u, y)$ also exhibits CARA with the same absolute risk aversion. Hence, if we show that DARA is maintained when adding background risk, the

results of Proposition 1 remain unchanged when applied to the class $\mathcal{U}^D$ of DARA utility functions allowing for additive background risk. To see this, suppose $u(\cdot) \in \mathcal{U}^D$. We will show that $v(\cdot; u, y) \in \mathcal{U}^D$ for any additive background risk y.

Let $A_u(w)$ and $A_v(w)$ denote the absolute risk aversion coefficients of $u(\cdot)$ and $v(\cdot; u, y)$, respectively, i.e.,

$$A_u(w) = -\frac{u''(w)}{u'(w)},$$

and

$$A_v(w) = -\frac{v''(w; u, y)}{v'(w; u, y)} = -\frac{\mathbf{E}[u''(w+y)]}{\mathbf{E}[u'(w+y)]}.$$

Since $u(\cdot) \in \mathcal{U}^D$, we know

$$A_u'(w) = \frac{[u''(w)]^2 - u'''(w)\, u'(w)}{[u'(w)]^2} \leq 0,$$

which implies

$$\left[u''(w)\right]^2 \leq u'''(w)\, u'(w). \tag{30}$$

Now,

$$\begin{aligned} A_v'(w) &= \frac{(\mathbf{E}[u''(w+y)])^2 - \mathbf{E}[u'''(w+y)]\,\mathbf{E}[u'(w+y)]}{(\mathbf{E}[u'(w+y)])^2} \\ &\leq \frac{(\mathbf{E}[u''(w+y)])^2 - \left(\mathbf{E}\left[\sqrt{u'''(w+y)\, u'(w+y)}\right]\right)^2}{(\mathbf{E}[u'(w+y)])^2}, \end{aligned} \tag{31}$$

where the inequality follows from the Cauchy-Schwarz inequality.[15] By (30), we have that for all realizations of y,

$$-u''(w+y) \leq \sqrt{u'''(w+y)\, u'(w+y)},$$

where the negative sign on the left-hand side results from the concavity of $u(\cdot)$. Taking expectation on both sides yields

$$-\mathbf{E}\left[u''(w+y)\right] \leq \mathbf{E}\left[\sqrt{u'''(w+y)\, u'(w+y)}\right],$$

which in turn implies

$$\left(\mathbf{E}\left[u''(w+y)\right]\right)^2 \leq \left(\mathbf{E}\left[\sqrt{u'''(w+y)\, u'(w+y)}\right]\right)^2. \tag{32}$$

[15] Note that by DARA, $u''' \geq 0$, implying that the term under the radical is non-negative.

Hence, from (31) and (32) we obtain $A_v'(w) \leq 0$. This completes the proof of the proposition. $\square$

Proof of Proposition 11. It is immediate that when $u(w) = \log w$, then a multiplicative background risk results in adding a constant to utility, and thus does not change the form of the utility function. Namely, log utility can be written in the form $v(\cdot; u, z) = \mathbf{E}[u(w(1+z))]$ for some $u \in \mathcal{U}^I$. By Remark (1) in Hart (2011, p. 637), if we can show that for any $u \in \mathcal{U}^I$ we have $v(\cdot; u, z) \in \mathcal{U}^I$, then the results of Proposition 2 remain unchanged when applied to the class $\mathcal{U}^I$, allowing for multiplicative background risk. To see this, let $u(\cdot) \in \mathcal{U}^I$, and consider a multiplicative background risk $1+z$.

Let $R_u(w)$ and $R_v(w)$ denote the relative risk aversion coefficients of $u(\cdot)$ and $v(\cdot; u, z)$, respectively, i.e.,

$$R_u(w) = -w\frac{u''(w)}{u'(w)},$$

and

$$R_v(w) = -w\frac{v''(w; u, z)}{v'(w; u, z)} = -w\frac{\mathbf{E}\left[(1+z)^2 u''(w(1+z))\right]}{\mathbf{E}\left[(1+z)u'(w(1+z))\right]}.$$

Since $u(\cdot) \in \mathcal{U}^I$, we know that

$$-(w(1+z))\frac{u''(w(1+z))}{u'(w(1+z))} \geq 1$$

for any z. Equivalently, after multiplying by $1+z \geq 0$,

$$-\left(w(1+z)^2\right)u''(w(1+z)) \geq (1+z)u'(w(1+z)).$$

Taking expectations on both sides with respect to z gives

$$-w\frac{\mathbf{E}\left[(1+z)^2 u''(w(1+z))\right]}{\mathbf{E}\left[(1+z)u'(w(1+z))\right]} \geq 1.$$

That is, $R_v(w) \geq 1$ as needed. $\square$

Appendix B

Table B1 Comparison of existing performance measures

This table reviews existing measures used for performance evaluation and their properties. We denote the portfolio return and the risk-free rate by r and r_f, respectively. The excess return of a portfolio is denoted by $g = r - r_f$. In addition, $\underline{g}$ stands for the lowest acceptable return level. For the manipulation-proof performance measure Θ (Goetzmann et al., 2007), we denote the time horizon and time length between observations by T, and Δt, and ρ is chosen to make holding the benchmark optimal for an uninformed investor. For each of the performance measures, we examine the following four properties: (1) whether it induces a complete order on the set of gambles; (2) whether it only depends on the distribution of portfolio returns to be evaluated; (3) whether it is consistent with SOSD whenever SOSD can be applied; and (4) whether it accounts for higher distribution moments in a manner consistent with asset pricing literature. We denote by "Y" a case where the measure satisfies a property, and by "N" a case where it does not.

Performance measure	Reference	Formula	Properties (1)	(2)	(3)	(4)
Sharpe ratio	Sharpe (1966)	$S = \frac{\mathbf{E}(g)}{[\mathbf{Var}(g)]^{1/2}}$	Y	Y	N	N
Jensen's alpha	Jensen (1968)	$\alpha = g - \sum_{k=1}^{K} \beta_k f_k$	Y	N[16]	N	N
Sortino downside risk measure	Sortino and van der Meer (1991)	$D = \frac{\mathbf{E}(g)}{(\mathbf{E}[\min^2(g-\underline{g},0)])^{1/2}}$	Y	Y	N	N
SVP upside potential measure	Sortino et al. (1999)	$U = \frac{\mathbf{E}[\max(g-\underline{g},0)]}{(\mathbf{E}[\min^2(g-\underline{g},0)])^{1/2}}$	Y	Y	N	N
Omega ratio	Keating and Shadwick (2002)	$\Omega = \frac{\int_{\underline{g}}^{+\infty} [1-F(g)]dg}{\int_{-\infty}^{\underline{g}} F(g)dg}$	Y	Y	N	N[17]
Kappa measure	Kaplan and Knowles (2004)	$K_n = \frac{\mathbf{E}(g)-\underline{g}}{\left[\int_{-\infty}^{\underline{g}} (\underline{g}-g)^n dF(g)\right]^{1/n}}$	Y	Y	N	N[18]
Manipulation-proof performance measure	Goetzmann et al. (2007)	$\Theta = \frac{1}{(1-\rho)\Delta t} \ln\left(\frac{1}{T}\sum_{t=1}^{T}\left[\left(\frac{1+r_t}{1+r_{ft}}\right)^{1-\rho}\right]\right)$	Y	N[19]	Y	Y
Generalized Sharpe ratio	Zakamouline and Koekebakker (2009)	$GSR = [-2\log(-\max_a \mathbf{E}[-\exp(-\lambda a g)])]^{1/2}$	Y	N[20]	Y[21]	Y

[16] The calculation of α depends on k benchmark portfolios.

[17] It is shown in Kaplan and Knowles (2004) that $\Omega = K_1 + 1$. Since K_n does not satisfy property (4), it follows that Ω does not either.

[18] For each n, K_n only accounts for the 1st and the nth moments.

[19] The calculation of Θ depends on the choice of a benchmark portfolio through ρ.

[20] The GSR is computed based on the assumption that the investor has CARA utility and zero initial wealth. It is therefore dependent on utility and wealth.

[21] In practice, GSR is usually estimated based on a truncated Taylor series. As a result, the estimated GSR is inconsistent with SOSD.

References

Aumann, R., Serrano, R., 2008. An economic index of riskiness. Journal of Political Economy 116, 810–836.

Bakshi, G., Chabi-Yo, F., Gao, X., 2011. Riskier times and asset returns. Unpublished working paper. University of Maryland, Ohio State University, University of Hong Kong.

Bali, T., Cakici, N., Chabi-Yo, F., 2011. A generalized measure of riskiness. Management Science 57, 1406–1423.

Banz, R., 1981. The relationship between return and market value of common stocks. Journal of Financial Economics 9, 3–18.

Barro, R., 2006. Rare disasters and asset markets in the twentieth century. Quarterly Journal of Economics 121, 823–866.

Barro, R., 2009. Rare disasters, asset prices, and welfare costs. American Economic Review 99, 243–264.

Barro, R., Jin, T., 2011. On the size distribution of macroeconomic disasters. Econometrica 79, 1567–1589.

Carhart, M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57–82.

Chang, E., Lewellen, W., 1984. Market timing and mutual fund investment performance. Journal of Business 57, 57–72.

Chen, H., Joslin, S., Tran, N., 2012. Rare disasters and risk sharing with heterogeneous beliefs. Review of Financial Studies 25, 2189–2224.

Eeckhoudt, L., Gollier, C., Schlesinger, H., 1996. Changes in background risk and risk taking behavior. Econometrica 64, 683–689.

Fama, E., French, K., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427–465.

Fama, E., French, K., 1993. Common risk factors in the returns on bonds and stocks. Journal of Financial Economics 33, 3–56.

Fama, E., French, K., 1996. Multifactor explanations of asset pricing anomalies. Journal of Finance 51, 55–84.

Foster, D., Hart, S., 2009. An operational measure of riskiness. Journal of Political Economy 117, 785–814.

Foster, D., Hart, S., 2013. A wealth-requirement axiomatization of riskiness. Theoretical Economics 8, 591–620.

Gabaix, X., 2008. Variable rare disasters: a tractable theory of ten puzzles in macro-finance. American Economic Review 98, 64–67.

Gabaix, X., 2012. Variable rare disasters: an exactly solved framework for ten puzzles in macro-finance. Quarterly Journal of Economics 127, 645–700.

Goetzmann, W., Ingersoll, J., Spiegel, M., Welch, I., 2007. Portfolio performance manipulation and manipulation-proof performance measures. Review of Financial Studies 20, 1503–1546.

Gourio, F., 2012. Disaster risk and business cycles. American Economic Review 102, 2734–2766.

Hadar, J., Russell, W., 1969. Rules for ordering uncertain prospects. American Economic Review 59, 25–34.

Hanoch, G., Levy, H., 1969. The efficiency analysis of choices involving risk. Review of Economic Studies 36, 335–346.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hart, S., 2011. Comparing risks by acceptance and rejection. Journal of Political Economy 119, 617–638.

Harvey, C., Siddique, A., 2000. Conditional skewness in asset pricing tests. Journal of Finance 55, 1263–1295.

Henriksson, R., 1984. Market timing and mutual fund performance: an empirical investigation. Journal of Business 57, 73–96.

Ippolito, R., 1989. Efficiency with costly information: a study of mutual fund performance, 1965–1984. Quarterly Journal of Economics 104, 1–23.

Jean, W., 1971. The extension of portfolio analysis to three or more parameters. Journal of Financial and Quantitative Analysis 6, 505–515.

Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.

Jensen, M., 1968. The performance of mutual funds in the period 1945–1964. Journal of Finance 23, 389–416.

Kane, A., 1982. Skewness preference and portfolio choice. Journal of Financial and Quantitative Analysis 17, 15–25.

Kaplan, P., Knowles, J., 2004. Kappa: a generalized downside risk-adjusted performance measure. Journal of Performance Measurement 8, 42–54.

Keating, C., Shadwick, W., 2002. A universal performance measure. Journal of Performance Measurement 6, 59–84.

Kraus, A., Litzenberger, R., 1976. Skewness preference and the valuation of risk assets. Journal of Finance 31, 1085–1100.

Lo, A., 2002. The statistics of Sharpe ratios. Financial Analysts Journal 58, 36–52.

Malkiel, B., 1995. Returns from investing in equity mutual funds 1971 to 1991. Journal of Finance 50, 549–572.

Moskowitz, T., Vissing-Jorgensen, A., 2002. The returns to entrepreneurial investment: a private equity premium puzzle? American Economic Review 92, 745–778.

Rosenberg, B., Reid, K., Lanstein, R., 1985. Persuasive evidence of market inefficiency. Journal of Portfolio Management 11, 9–17.

Rothschild, M., Stiglitz, J., 1970. Increasing risk I: a definition. Journal of Economic Theory 2, 225–243.

Rubinstein, M., 1973. The fundamental theorem of parameter-preference security valuation. Journal of Financial and Quantitative Analysis 8, 61–69.

Sharpe, W., 1966. Mutual fund performance. Journal of Business 39, 119–138.

Sortino, F., van der Meer, R., 1991. Downside risk. Journal of Portfolio Management 17, 27–31.

Sortino, F., van der Meer, R., Plantinga, A., 1999. The Dutch triangle. Journal of Portfolio Management 26, 50–57.

Wachter, J., 2013. Can time-varying risk of rare disasters explain aggregate stock market volatility? Journal of Finance 68, 987–1035.

Wermers, R., 2000. Mutual fund performance: an empirical decomposition into stock-picking talent, style, transactions costs, and expenses. Journal of Finance 55, 1655–1703.

Zakamouline, V., Koekebakker, S., 2009. Portfolio performance evaluation with generalized Sharpe ratios: beyond the mean and variance. Journal of Banking and Finance 33, 1242–1254.

Table 1 Motivating example I

This table provides an example in which the Sharpe ratio fails to capture the high-moment and rare disaster properties of the distribution of gambles, leading to an inconsistency with SOSD. Two gambles g_1 and g_2 are being compared, with the gamble value and the corresponding probability and cumulative distribution function (CDF) reported. The table also reports the various moments of the two gambles, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. The table also reports the Sharpe ratios of the gambles, denoted by S.

	g_1			g_2		
	Value	Probability	CDF	Value	Probability	CDF
	-10	0.001	0.001	-1	0.001	0.001
	1	0.999	1	1	0.9	0.901
				4	0.099	1
μ_1		0.989			1.295	
m_2		0.121			0.808	
m_3		-1.327			1.924	
m_4		14.583			5.335	
S		2.845			1.441	

Table 2 Motivating example II

This table provides an example in which the Sharpe ratio fails to capture the high-moment properties of the distribution of gambles, leading to an inconsistency with SOSD. Two gambles g_1 and g_2 are being compared, with the gamble value and the corresponding probability and CDF reported. The table also reports the various moments of the two gambles, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. The table also reports the Sharpe ratios of the gambles, denoted by S.

	g_1			g_2		
	Value	Probability	CDF	Value	Probability	CDF
	-1	0.001	0.001	-1	0.001	0.001
	1	0.999	1	1	0.9	0.901
				4	0.099	1
μ_1		0.998			1.295	
m_2		0.004			0.808	
m_3		-0.008			1.924	
m_4		0.016			5.335	
S		15.788			1.441	

Table 3 Attractiveness of anomalies

This table compares the attractiveness of investment in the market excess return ($mktrf$), and the small-minus-big (smb), high-minus-low (hml), and up-minus-down (umd) portfolios, based on the monthly returns in percentage points for the period January 1962 to December 2009. Panel A reports GMM estimates of various moments of the portfolio returns, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. Panel B reports the Sharpe ratio (S) and the Aumann-Serrano and Foster-Hart performance measures (P^{AS} and P^{FH}) for different investment portfolios. This panel also reports the absolute elasticities of the P^{AS} (P^{FH}) measures with respect to the normalized moments, denoted as η_1^{AS}, ζ_2^{AS}, ζ_3^{AS}, and ζ_4^{AS} (η_1^{FH}, ζ_2^{FH}, ζ_3^{FH}, and ζ_4^{FH}) (see Section 4.2). For the hml portfolio, the absolute elasticities of the P^{FH} measure cannot be computed because $-1 < P^{FH}(g)g \leq 1$ and $-1 < P^{FH}(g)g \leq 1 + 2P^{FH}(g)\mu_1(g)$ are violated, and hence, the Taylor expansions needed for the calculation of the absolute elasticities do not converge. Further, this panel reports the certainty equivalent (CE) of a risk-averse investor with CRRA utility and a risk aversion coefficient of 3, 5, or 10, assuming that the initial wealth and the investment scale are equal. In both panels, columns 1–4 report estimates for the four portfolios, respectively, whereas the difference test results are shown in columns 5–10. The t-statistics are reported in the parentheses below the corresponding estimates. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Panel A: Various moments

	(1) $mktrf$	(2) smb	(3) hml	(4) umd	(5) $mktrf - smb$	(6) $mktrf - hml$	(7) $mktrf - umd$	(8) $smb - hml$	(9) $smb - umd$	(10) $hml - umd$
μ_1	0.4059	0.2298	0.4391	0.7257	0.1761	-0.0332	-0.3198	-0.2093	-0.4959	-0.2866
	(2.15)**	(1.75)*	(3.62)***	(4.02)***	(0.91)	(-0.13)	(-1.14)	(-1.05)	(-2.20)**	(-1.23)
m_2	20.4828	9.9792	8.4912	18.7827	10.5036	11.9917	1.7001	1.4880	-8.8035	-10.2915
	(12.02)***	(8.74)***	(11.42)***	(6.71)***	(5.53)***	(7.24)***	(0.57)	(1.44)	(-3.28)***	(-3.85)***
m_3	-50.9774	16.6665	-0.6708	-116.1113	-67.6439	-50.3066	65.1339	17.3373	132.7778	115.4405
	(-2.02)**	(0.87)	(-0.11)	(-1.46)	(-2.19)**	(-1.91)*	(0.77)	(0.76)	(1.66)*	(1.44)
m_4	2,092.19	850.61	390.76	4,862.86	1,241.58	1,701.42	-2,770.67	459.85	-4,012.25	-4,472.10
	(3.45)***	(2.03)**	(4.27)***	(1.73)*	(1.71)*	(2.81)***	(-0.97)	(1.21)	(-1.42)	(-1.59)

Panel B: Performance comparison

	(1) $mktrf$	(2) smb	(3) hml	(4) umd	(5) $mktrf - smb$	(6) $mktrf - hml$	(7) $mktrf - umd$	(8) $smb - hml$	(9) $smb - umd$	(10) $hml - umd$
S	0.0897	0.0728	0.1507	0.1675	0.0169	-0.0610	-0.0778	-0.0779	-0.0947	-0.0168
					(0.35)	(-0.89)	(-1.13)	(-1.19)	(-1.46)	(-0.25)
P^{AS}	0.0382	0.0468	0.1014	0.0630	-0.0086	-0.0632	-0.0248	-0.0545	-0.0162	0.0384
					(-0.32)	(-1.63)	(-0.86)	(-1.22)	(-0.47)	(1.03)
η_1^{AS}	0.9602	1.0081	0.9638	0.7716						
ζ_2^{AS}	1.8263	2.0320	1.8101	1.2041						
ζ_3^{AS}	0.0869	0.0795	0.0072	0.2345						
ζ_4^{AS}	0.0454	0.0634	0.1427	0.2063						
P^{FH}	0.0347	0.0449	0.0800	0.0288	-0.0101	-0.0452	0.0059	-0.0351	0.0160	0.0512
					(-0.45)	(-2.86)***	(0.43)	(-1.42)	(0.71)	(9.46)***
η_1^{FH}	0.7510	0.8508		0.0022						
ζ_2^{FH}	1.2827	1.6264		0.0016						
ζ_3^{FH}	0.1094	0.1206		0.0003						
ζ_4^{FH}	0.1538	0.2733		0.0003						
CE ($\gamma = 3$)	0.0854	0.0822	0.3115	0.4071	0.0032	-0.2261	-0.3218	-0.2293	-0.3250	-0.0957
					(0.02)	(-0.85)	(-1.02)	(-1.15)	(-1.25)	(-0.36)
CE ($\gamma = 5$)	-0.1440	-0.0161	0.2251	0.1379	-0.1279	-0.3692	-0.2820	-0.2413	-0.1541	0.0872
					(-0.61)	(-1.34)	(-0.76)	(-1.19)	(-0.49)	(0.27)
CE ($\gamma = 10$)	-0.8013	-0.2684	0.0024	-1.0615	-0.5329	-0.8037	0.2602	-0.2708	0.7931	1.0639
					(-1.93)*	(-2.38)**	(0.29)	(-1.24)	(0.93)	(1.25)

Table 4 Optimal combinations of anomalies

This table illustrates the pairwise optimal convex combinations of the market excess return ($mktrf$), and the small-minus-big (smb), high-minus-low (hml), and up-minus-down (umd) portfolios using monthly returns in percentage points for the period January 1962 to December 2009. Panels A and B report the results for the P^{AS} and P^{FH} measures, respectively. For each pair of portfolios, the table shows the maximal P^{AS} and P^{FH} values delivered by the optimal convex combination, with the associated weights assigned to the two portfolios reported below in parentheses.

Panel A: P^{AS} measure				
	$mktrf$	smb	hml	umd
$mktrf$		0.0659 (0.35,0.65)	0.1941 (0.33,0.67)	0.1232 (0.46,0.54)
smb			0.2006 (0.43,0.57)	0.1304 (0.57,0.43)
hml				0.1933 (0.68,0.32)
umd				

Panel B: P^{FH} measure				
	$mktrf$	smb	hml	umd
$mktrf$		0.0606 (0.31,0.69)	0.1103 (0.38,0.62)	0.0685 (0.44,0.56)
smb			0.1540 (0.54,0.46)	0.0802 (0.56,0.44)
hml				0.1143 (0.84,0.16)
umd				

Table 5 High-moment and rare disaster properties of momentum portfolio

This table illustrates the high-moment and rare disaster properties of the momentum (*umd*) portfolio, based on the monthly *umd* returns in percentage points for the period January 1962 to December 2009. The results for the *umd* portfolio are reported in column 1. In comparison, we draw one million samples of the same size from a simulated normal distribution with mean and variance equal to those of the sample momentum returns. The results for the simulated distribution are reported in column 2. In addition, difference test results are reported in column 3. Panel A shows estimates of various moments of the portfolio returns, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. Panel B reports the Sharpe ratio (S) and the Aumann-Serrano and Foster-Hart performance indices (P^{AS} and P^{FH}). Further, Panel B also reports the certainty equivalent (CE) of a risk-averse investor with CRRA utility and a risk aversion coefficient of 3, 5, or 10, assuming that the initial wealth and the investment scale are equal. Bootstrapped t-statistics are reported in the parentheses below the corresponding estimates. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Panel A: Various moments

	(1) *umd*	(2) Simulated	(3) *umd*-Simulated
μ_1	0.7257	0.7257	0.0001
	(4.02)***	(96.32)***	(0.00)
m_2	18.7827	18.7504	0.0323
	(6.73)***	(407.23)***	(0.01)
m_3	-116.1113	-0.0026	-116.1086
	(-1.47)	(-0.01)	(-1.47)
m_4	4,862.86	1,054.80	3,808.06
	(1.74)*	(176.42)***	(1.36)

Panel B: Performance comparison

	(1) *umd*	(2) Simulated	(3) *umd*-Simulated
S	0.1675	0.1678	-0.0003
			(-0.01)
P^{AS}	0.0630	0.0777	-0.0147
			(-0.63)
P^{FH}	0.0288	0.0674	-0.0386
			(-2.94)***
CE ($\gamma = 3$)	0.4071	0.4450	-0.0379
			(-0.17)
CE ($\gamma = 5$)	0.1379	0.2566	-0.1186
			(-0.43)
CE ($\gamma = 10$)	-1.0615	-0.2192	-0.8424
			(-1.04)

Table 6 Private vs. public equity

This table compares the performance of private (column 1) and public (column 2) equity. The private equity performance is evaluated based on entrepreneur-level returns constructed from the 2004 Survey of Consumer Finance (SCF). This sample is conditional on survival, yielding an upward bias in the evaluation of private equity performance. The public equity returns are proxied by returns to the CRSP value-weighted market index. The difference test results are reported in column 3. Panel A shows GMM estimates of various moments of the equity returns, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. Panel B reports the Sharpe ratio (S) and the Aumann-Serrano and Foster-Hart performance indices (P^{AS} and P^{FH}). Further, Panel B also reports the certainty equivalent (CE) of a risk-averse investor with CRRA utility and a risk aversion coefficient of 3 or 5, assuming that the initial wealth and the investment scale are equal. We are not able to compute the certainty equivalent for a risk aversion coefficient of 10 since then, investing in private equity is almost equivalent to losing the entire initial wealth for sure, rendering the calculations computationally impossible. The t-statistics are reported in the parentheses below the corresponding estimates. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Panel A: Various moments

	(1) Private	(2) Public	(3) Private-Public
μ_1	196.5432	5.5895	190.9537
	(5.26)***	(92.12)***	(5.11)***
m_2	6.1506×10^6	16.2216	6.1505×10^6
	(1.90)*	(23.29)***	(1.90)*
m_3	5.1344×10^{11}	73.0088	5.1344×10^{11}
	(1.75)*	(8.58)***	(1.75)*
m_4	4.6177×10^{16}	2,400.33	4.6177×10^{16}
	(1.74)*	(15.69)***	(1.74)*

Panel B: Performance comparison

	(1) Private	(2) Public	(3) Private-Public
S	0.0793	1.3878	-1.3086
			(-45.96)***
P^{AS}	0.0549	0.7971	-0.7422
			(-38.18)***
P^{FH}	0.0100	0.2583	-0.2483
			(-4,293,137)***
CE ($\gamma = 3$)	-70.3273	5.3536	-75.6808
			(-12.04)***
CE ($\gamma = 5$)	-94.4017	5.2129	-99.6146
			(-156.16)***

Table 7 Active vs. passive funds

This table compares the performance of actively managed (column 1) and index (column 2) funds, based on the CRSP mutual fund return data for the time period January 1991 to December 2009. The difference test results are reported in column 3. Panel A shows GMM estimates of various moments of the fund returns, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. Panel B reports the Sharpe ratio (S) and the Aumann-Serrano and Foster-Hart performance indices (P^{AS} and P^{FH}). Further, Panel B also reports the certainty equivalent (CE) of a risk-averse investor with CRRA utility and a risk aversion coefficient of 3, 5, or 10, assuming that the initial wealth and the investment scale are equal. The t-statistics are reported in the parentheses below the corresponding estimates. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Panel A: Various moments

	(1) Active	(2) Passive	(3) Active-Passive
μ_1	0.5210	0.5409	-0.0199
	(1.79)*	(1.88)*	(-0.30)
m_2	19.2097	18.9356	0.2741
	(7.90)***	(8.08)***	(0.44)
m_3	-67.3233	-58.9134	-8.4099
	(-2.24)**	(-2.17)**	(-1.38)
m_4	1,718.30	1,610.39	107.91
	(2.81)***	(2.96)***	(0.97)

Panel B: Performance comparison

	(1) Active	(2) Passive	(3) Active-Passive
S	0.1189	0.1243	-0.0054
			(-0.37)
P^{AS}	0.0508	0.0537	-0.0029
			(-0.49)
P^{FH}	0.0441	0.0465	-0.024
			(-0.64)
CE ($\gamma = 3$)	0.2175	0.2435	-0.0260
			(-0.40)
CE ($\gamma = 5$)	-0.0014	0.0307	-0.0321
			(-0.50)
CE ($\gamma = 10$)	-0.6240	-0.5665	-0.0575
			(-0.81)

Table 8 Mutual fund selection based on performance measures

This table reports the performance of investments following monthly rebalancing strategies based on the P^{AS} measure (column 1), the P^{FH} measure (column 2), and the Sharpe ratio (column 3) for the period January 1967 to December 2009. For comparison, the table also reports the performance of investing in the market, represented by the CRSP value-weighted index (column 4) for the same sample period. Columns 5–10 show the difference test results. Panel A reports GMM estimates of various moments of the portfolio returns, where μ_1 denotes the first moment, and m_2, m_3, and m_4 stand for the second, third, and fourth central moments, respectively. Panel B reports the Sharpe ratio (S) and the Aumann-Serrano and Foster-Hart performance measures (P^{AS} and P^{FH}). Further, this panel also reports the certainty equivalent of a risk-averse investor with CRRA utility and risk aversion coefficients of 3, 5, and 10, which are typical in asset pricing calibration exercises. We assume that the initial wealth and the investment scale are equal. The t-statistics are reported in the parentheses below the corresponding estimates. Asterisks denote statistical significance at the 1% (***), 5% (**), and 10% (*) levels.

Panel A: Various moments

	(1) AS	(2) FH	(3) S	(4) MKT	(5) AS-FH	(6) AS-S	(7) AS-MKT	(8) FH-S	(9) FH-MKT	(10) S-MKT
μ_1	0.4961	0.4671	0.5383	0.4269	0.0290	-0.0422	0.0692	-0.0712	0.0402	0.1114
	(2.80)***	(2.68)***	(2.66)***	(2.09)**	(1.89)*	(-0.96)	(1.13)	(-1.32)	(0.62)	(1.77)*
m_2	16.1544	15.6887	21.2108	21.4319	0.4657	-5.0564	-5.2774	-5.5221	-5.7432	-0.2211
	(11.31)***	(11.06)***	(11.89)***	(11.46)***	(3.55)***	(-9.92)***	(-6.82)***	(-9.67)***	(-7.49)***	(-0.27)
m_3	-37.2520	-36.9873	-59.1041	-55.5719	-0.2647	21.8521	18.3199	22.1168	18.5846	-3.5321
	(-2.23)**	(-2.17)**	(-2.62)***	(-1.99)**	(-0.18)	(2.85)***	(1.29)	(2.76)***	(1.38)	(-0.25)
m_4	1,312.89	1,284.98	2,090.75	2,264.29	27.91	-777.86	-951.40	-805.77	-979.31	-173.55
	(4.12)***	(3.92)***	(4.43)***	(3.35)***	(0.96)	(-4.29)***	(-2.34)**	(-4.36)***	(-2.53)**	(-0.48)

Panel B: Performance comparison

	(1) AS	(2) FH	(3) S	(4) MKT	(5) AS-FH	(6) AS-S	(7) AS-MKT	(8) FH-S	(9) FH-MKT	(10) S-MKT
S	0.1234	0.1179	0.1169	0.0922	0.0055	0.0066	0.0312	0.0011	0.0257	0.0247
					(1.42)	(0.77)	(2.48)**	(0.10)	(1.92)*	(1.81)*
P^{AS}	0.0582	0.0565	0.0482	0.0384	0.0018	0.0100	0.0199	0.0082	0.0181	0.0099
					(0.97)	(2.12)**	(3.16)***	(1.43)	(2.69)***	(1.74)*
P^{FH}	0.0497	0.0484	0.0424	0.0347	0.0013	0.0074	0.0150	0.0061	0.0137	0.0076
					(1.17)	(2.95)***	(3.89)***	(1.91)*	(3.47)***	(1.78)*
CE ($\gamma = 3$)	0.2450	0.2229	0.2060	0.0910	0.0220	0.0390	0.1540	0.0170	0.1320	0.1150
					(1.40)	(0.83)	(2.44)**	(0.30)	(1.98)**	(1.77)*
CE ($\gamma = 5$)	0.0672	0.0498	-0.0318	-0.1501	0.0174	0.0990	0.2173	0.0816	0.1998	0.1182
					(1.08)	(1.95)*	(3.17)***	(1.36)	(2.80)***	(1.68)*
CE ($\gamma = 10$)	-0.4266	-0.4334	-0.7035	-0.8434	0.0068	0.2769	0.4168	0.2701	0.4100	0.1399
					(0.37)	(4.02)***	(3.63)***	(3.50)***	(3.64)***	(1.24)

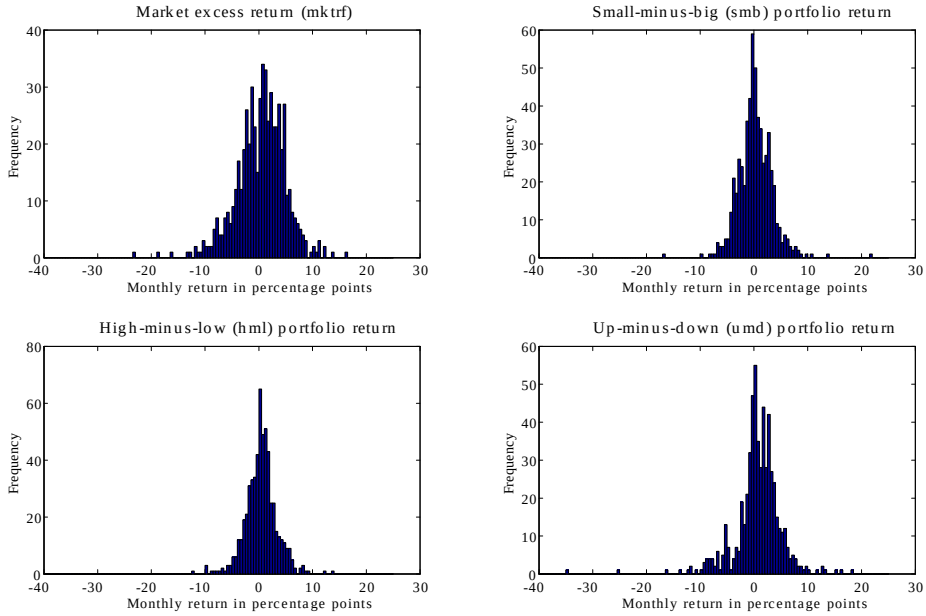

Fig. 1. Distributions of anomaly returns.

This figure shows the histograms of the monthly market excess return ($mktrf$), and the monthly returns of the small-minus-big (smb), high-minus-low (hml), and up-minus-down (umd) portfolios for the period January 1962 to December 2009.

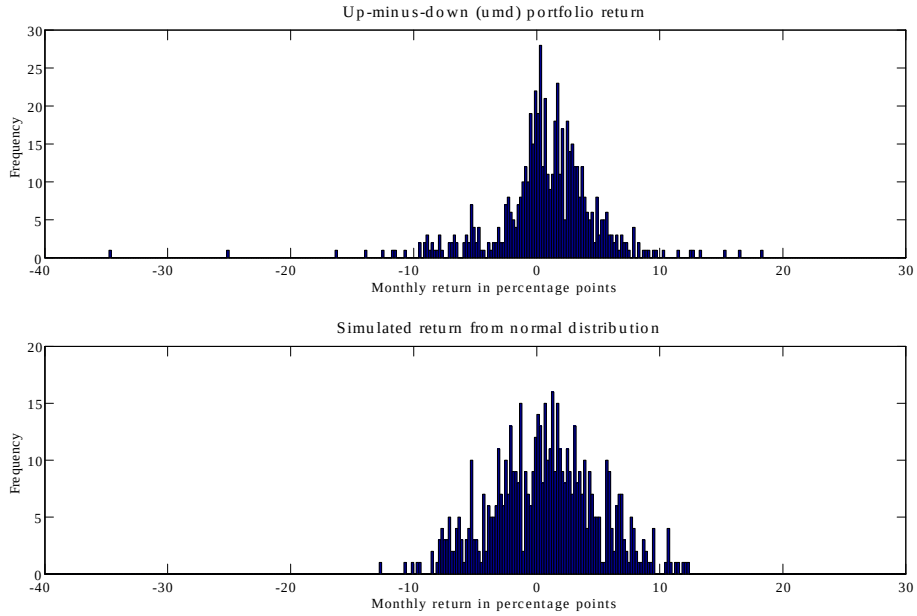

Fig. 2. Distributions of momentum and simulated normal returns.

This figure compares the histogram of the monthly returns of the momentum (*umd*) portfolio against the histogram of a simulated normal distribution with the same mean and variance for the period January 1962 to December 2009.

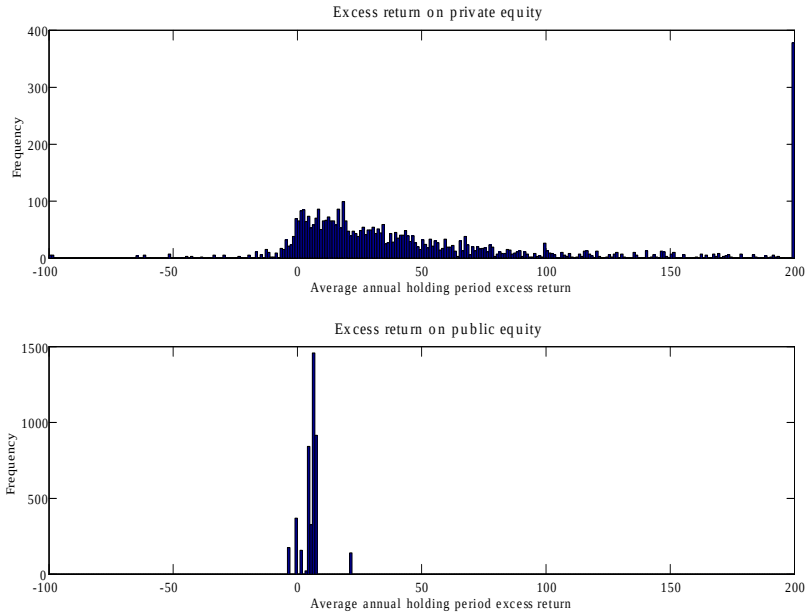

Fig. 3. Distributions of private and public equity returns.

This figure compares the histograms of the average annual holding period excess returns from private versus public equity investments. The private equity returns are estimated at the entrepreneur level based on the 2004 Survey of Consumer Finance (SCF). This sample is conditional on survival, and the returns are winsorized at 200%. The public equity returns are proxied by returns to the CRSP value-weighted market index for each household during the same time period as its private equity holdings.