A Probabilistic Risk-to-Reward Measure for Evaluating the Performance of Financial Securities

Phil Maguire, Philippe Moser, Jack McDonnell, Robert Kelly, Simon Fuller, Rebecca Maguire

***Abstract*—Existing risk-to-reward measures, such as the Sharpe ratio [1] or M2 [2], are based on the idea of quantifying the excess return per unit of deviation in an investment. In this preliminary article we introduce a new probabilistic measure for evaluating investment performance. Randomness Deficiency Coefficient (RDC) expresses the likelihood that the observed excess return of an investment has been generated by chance. Some of the advantages of RDC over existing measures are that it can be used with small historical datasets, is time-frame independent, and can be easily adjusted to take into account the familywise error rate which results from selection bias. We argue that RDC captures the fundamental relationship between risk and reward and prove that it converges with Sharpe's ratio.**

I. Introduction

It is important for investors and financial analysts to be able to measure and compare the performance of different investments. One simple approach is to compare the level of return earned over a given period. The financial industry continues to rely heavily on simple return for evaluating fund performance, with investment managers often benchmarked against the returns generated by an unmanaged market, or a capitalization-weighted portfolio [2]. The problem with this approach is that it ignores the risk involved in generating the return, and hence fails to communicate the significance of that return.

Consider Figure I below in which the returns of two portfolios are presented. Both yield an identical return of 12%, yet portfolio B seems intuitively less appealing. Because it has a higher level of volatility, the returns earned in this case are more likely to represent random variance, as opposed to underlying performance: they do not strongly challenge the hypothesis that the investment is simply following a random walk. In the case of portfolio A, the lower volatility suggests that the return is less likely to have been produced by chance, and thus more likely to persist into the future. The lower the variance in the returns of an investment, the more unlikely it is that the performance emerged by chance alone.

Rebecca Maguire is with the School of Business, National College of Ireland. All other authors are with the Department of Computer Science, National University of Ireland, Maynooth, Ireland. Correspondence should be addressed to Phil Maguire (phone: 353-1-7086082; fax: 353-1-7083848; e-mail: pmaguire@cs.nuim.ie).

This work was supported in part by the Irish Research Council for Science, Engineering and Technology through an Empower Research Fellowship awarded to Phil Maguire and an Embark Postgraduate Scholarship awarded to Simon Fuller.

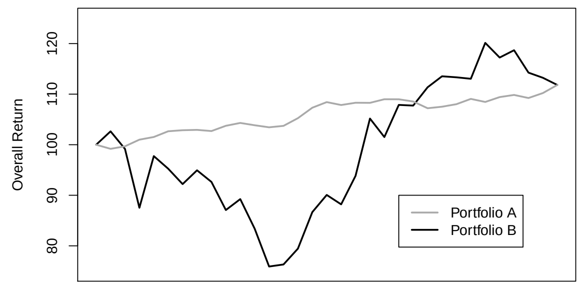

Fig. 1. Two portfolios yielding a return of 12%.

In sum, reliance on simple return can be misleading. What is needed for comparing risky investments is a means of adjusting returns for the level of risk involved.

A. Measures of Risk-to-Reward

The most commonly used measure of risk-adjusted return is the Sharpe ratio, which takes into account the ratio between the reward and the variability in value to which the investor was exposed. Here, the risk-free rate (the amount that could have been earned without holding any risk) is subtracted from the total returns, and the remainder is divided by the portfolio's standard deviation, thus effectively providing a measure of reward per unit of risk [1].

The Sharpe ratio can be used to evaluate both historical and predicted performance, with Sharpe [3] defining both ex post and ex ante versions of his ratio for these different applications. Given an investment yielding an expected return R with standard deviation σ and expected risk-free return R_f, then Sharpe's ex ante ratio is given by

$$S = \frac{R - R_f}{\sigma}.$$

Even though the ex ante version might be justified on the basis of predicted relationships, Sharpe states that, because R and σ are unobservable, they are inevitably estimated using historic data [3]. Given a sample of historical returns $R_1, \ldots, R_n$ and a constant risk free rate R_f the sampled components of the Sharpe ratio are given by

$$\hat{R} = \frac{1}{n}\sum_{i=1}^{n}(R_i - R_f) \quad \text{and} \quad \hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(R_i - R_f - \hat{R})^2.$$

The Sharpe ratio always refers to the differential between two portfolios, in this case that between the risky investment and the risk-free investment. The subtracted return reflects the short position which must be taken to finance the acquisition. To yield a positive Sharpe ratio, an investment must provide a greater return than the cash or loan which is used to fund it [4].

One problem with Sharpe's ratio is that the value it yields is dimensionless, making its significance difficult to interpret, particularly for negative values. Modigliani risk-adjusted performance or M2 is a refinement of the Sharpe ratio which resolves this issue by expressing the excess return earned over a standardized risky benchmark such as the market. Writing σ_B for the standard deviation of the excess returns of some benchmark portfolio (e.g. the market), and $\overline{R_M}$ for the average Market Return for the given period then M2 is given by

$$M2 = SR \cdot \sigma_B - \overline{R_M}.$$

Essentially, M2 adjusts the level of risk of the investment to match that of the benchmark and then computes the excess or shortfall return delivered by the investment over the benchmark at that standardized level of risk. This measure has the advantage of outputting a percentage, which is more intuitive to interpret for investors. However, it is worth noting that any rankings of investments using M2 will be identical to the rankings computed using Sharpe's ratio, as M2 is simply the Sharpe ratio adjusted by a constant for the purpose of enhancing interpretability. As a result, M2 inherits the limitations of the Sharpe ratio, which we detail in the following section.

B. Limitations of Sharpe's ratio

Sharpe's ratio seeks to provide a measure of reward to variability by expressing return in terms of deviation, thus allowing risky securities with different volatility profiles to be compared against each other. Other related measures, such as M2, which like the Sharpe ratio takes into account idiosynractic risk, or the Treynor ratio [5] and Jensen's alpha [6], which focus on systematic risk, express the reward to variability profile of a security relative to that of a risky benchmark such as the overall market.

In all cases, these measures are limited to comparisons between risky investments. The Sharpe ratio, for example, can say nothing about whether exposure to the risk of the stock market is better than investing in risk-free short-term Treasury Bills. Any risk free investment will have zero variability, meaning that its Sharpe ratio is undefined. Thus, what the Sharpe ratio provides is not a genuine measure of the relationship between risk and reward but, rather, a heuristic for ranking risky investments relative to each other, which does not express the extent to which the risk is justified in the first instance.

According to Sharpe the ratio is simply "a convenient summary of two important aspects in any strategy..." In recognition of the fact that return is good and variability is bad, the Sharpe ratio divides one by the other. However, this approach requires "a substantial set of assumptions for justification...[which in practice]...are, at best, likely to hold only approximately" [3]. First, the use of Sharpe's ratio assumes that the returns of all securities follow the same distribution, typically assumed to be the normal distribution. Second, it assumes and that the true mean and standard deviation can be identified with precision based on sampled data.

The smaller the volume of historical data available, the less satisfactorily these requirements can be met. For example, Bailey and de Pardo [7] have shown that a typical hedge fund's track record exhibits negative skewness and positive excess kurtosis, which has the effect of inflating its Sharpe ratio for smaller datasets. They conclude that the ratio can only be used to evaluate the performance of an investment when the size of the dataset exceeds a certain threshold, and provide a method for establishing the number of samples needed for computing Sharpe's ratio with a given confidence level.

Another problem is that Sharpe's ratio is time-frame dependent. The larger the period intervals over which returns are reported for a given duration, the lower the standard deviation, and the higher the resulting Sharpe ratio. Thus, the same investment will yield different ratios depending on whether returns are calculated daily, weekly or monthly, opening up the possibility of manipulation [8]. In the following section we present a probabilistic measure which captures the underlying relationship between risk and reward, and addresses these limitations of Sharpe's ratio.

II. Randomness Deficiency Coefficient

Sharpe's ratio is based on the idea of adjusting returns for the level of risk involved. For example, if investment A has twice the variance of investment B then we should halve the returns of investment A before comparing them, to standardize the level of volatility. But how do we compare investment A to a risk-free investment? How can we tell if the burden of risk is justified in the first instance? Our approach is to consider the underlying relationship between risk and reward. Anybody walking into a casino has the opportunity to trade risk for potential rewards. For example, a \$1 note can be exchanged for a 10% chance of holding \$10 or a 1% chance of holding \$100 (assuming the casino offers fair odds). Although such gambling strategies might provide temporary gains, the law of large numbers states that, over the longer run, the average return will converge on the expected value of 0.

Accordingly, the question that should be asked when a security delivers a particular return is whether that return is over and above that which could have been achieved by simply gambling: that is, the extent to which the returns provide evidence of performance over luck. No matter how consistent and extensive the levels of return, one can never be absolutely certain that they are not simply the result of luck alone. However, it is possible to measure the extent to which the observed pattern of returns deviates from that which would be expected given a random walk centered about some benchmark level of return. In the following sections we show how the randomness deficiency of a series of returns can be quantified by applying either a known probability distribution or, when there is insufficient data for identifying distribution parameters, a bootstrapping technique.

A. RDC: Randomness Deficiency Coefficient of returns

The Randomness Deficiency Coefficient (RDC) of an investment is the inverse probability that the observed level of excess returns should emerge by chance given a random walk centered on some chosen value (e.g. the risk-free rate, market return or other selected benchmark). To evaluate this probability we first need to specify the set of potential events from which the observed events are drawn.

Assuming the samples follow a known distribution[1], then we can use a probability density function (PDF) to quickly compute the probability that a sample of size n drawn randomly from this distribution will exceed the sample mean by the observed margin.

More precisely, given $n+1$ consecutive values $a_1, a_2, \dots, a_{n+1}$ of security t, we consider the log of the returns $r_1 = \log a_2/a_1, r_2 = \log a_3/a_2, \dots, r_n = a_{n+1}/a_n$ of t. Our skeptical hypothesis H_0^t says in essence that t's returns cannot exceed some a priori bound μ_B, where μ_B can be chosen to be e.g. the risk-free rate R_f, the market return, etc, and that t's returns are normally distributed, i.e. $H_0^t:\ R_1, R_2 \dots, R_n$ are i.i.d. normally distributed random variables with mean μ_t and standard deviation σ_t, where $\mu_t \leq \mu_B$. Thus the sample mean $\bar{R}$ is $\mathrm{Norm}(\mu_t, \sigma_t/\sqrt{n})$. Denote by m_t and s_t the sample mean and sample standard deviation of t, computed from n observed returns of t.

Definition 1: The RDC of security t is the inverse of the probability that the security's return is at least m_t under H_0^t. To make the value more inteligible, we rescale RDC to take values in $(-\infty, -1] \cup [1, \infty)$ (by considering cases $m_t < \mu_t$ and $m_t \geq \mu_t$ separately, i.e. given

$$H_0^t : R_i \sim N(\mu_t, \sigma_t),\ i = 1, 2, \dots, n$$

If $m_t \geq \mu_t$ then

$$RDC(t) = [\Pr(\bar{R} \geq m_t |\ H_0^t \text{ and } m_t \geq \mu_t)]^{-1}$$

If $m_t < \mu_t$ then we fix RDC to be negative i.e.

$$RDC(t) = -[\Pr(\bar{R} \leq m_t |\ H_0^t \text{ and } m_t < \mu_t)]^{-1}$$

Under H_0^t, $\bar{R}$ is $\mathrm{Norm}(\mu_t, \sigma_t/\sqrt{n})$. If we substitute the sample standard deviation s_t for σ_t then $\sqrt{n}(\bar{R} - \mu_t)/\sigma_t$ is t-distributed with $n-1$ degrees of freedom.

When μ_t is unknown, we assume that $\mu_t = \mu_B$, i.e. the security does its best.

We write $pdfRDC$ to specify the PDF-definition is used to compute the RDC.

[1] See our recent work [?] for a more general framework

B. Bootstrapping method

For small sample sizes, where the volume of historical data is too limited to reliably infer a specific probability distribution in the returns, it may be desirable to use a bootstrapping method to compute RDC. We here describe such a method.

Adhering to the null hypothesis that returns are produced by chance, we assume that for each sampled change in the value of an investment, an equal change in the opposite direction was equally likely. For each sample there are thus two possibilities to choose from, yielding 2^n possible sequences of events, where n is the number of samples. The RDC is then simply the inverse of the proportion of cases where the cumulative return of a randomized sequence of such events matches (or exceeds) the observed return. This value reflects the level of confidence with which the skeptical random hypothesis can be rejected.

For example, suppose that a mutual fund posts quarterly returns of +2.3%, +1.2%, -0.4% and +1.6% over the risk free rate. The total return, using arithmetic means, is 4.7%. Then the full set of possible scenarios involving these changes is given as shown in Table 1. Effectively, the small sample is bootstrapped by including the balancing values which would be expected under the null hypothesis. Although we have not done so in the example below, returns should be logged to faciliate the calculation of geometric as opposed to arithmetic means.

Q1	Q2	Q3	Q4	SUM	>= 4.7%?
+2.3%	+1.2%	+0.4%	+1.60%	+5.50%	YES
+2.3%	+1.2%	+0.4%	-1.60%	+2.30%	NO
+2.3%	+1.2%	-0.4%	-1.60%	+1.50%	NO
+2.3%	+1.2%	-0.4%	+1.60%	+4.70%	YES
+2.3%	-1.2%	+0.4%	+1.60%	+3.10%	NO
+2.3%	-1.2%	+0.4%	-1.60%	-0.10%	NO
+2.3%	-1.2%	-0.4%	+1.60%	+2.30%	NO
+2.3%	-1.2%	-0.4%	-1.60%	-0.90%	NO
-2.3%	+1.2%	+0.4%	+1.60%	+0.90%	NO
-2.3%	+1.2%	+0.4%	-1.60%	-2.30%	NO
-2.3%	+1.2%	-0.4%	-1.60%	-3.10%	NO
-2.3%	+1.2%	-0.4%	+1.60%	+0.10%	NO
-2.3%	-1.2%	+0.4%	+1.60%	-1.50%	NO
-2.3%	-1.2%	+0.4%	-1.60%	-4.70%	NO
-2.3%	-1.2%	-0.4%	+1.60%	-2.30%	NO
-2.3%	-1.2%	-0.4%	-1.60%	-5.50%	NO

TABLE I
FULL SET OF SCENARIOS BASED ON A SERIES OF FOUR RETURNS

The RDC is the total number of positive scenarios divided by the number of randomized scenarios which match the observed return, which in this case is 8/2 or 4. Because we are evaluating a one-tailed hypothesis we restrict the analysis to the set of positive scenarios: given that a positive return is observed, how strongly can we reject the hypothesis that it derives from the set of randomly sampled positive returns? An RDC of 4 means that, on average, one would need to select the best from a group of four randomly generated time series with positive returns to match the performance of this security (i.e. one quarter of randomly generated time series with overall positive return are at least as good as the one under consideration). In cases where n is too large to evaluate the full set of possible scenarios as above then we can use a Monte Carlo algorithm to sample a random subset of scenarios on which RDC can be computed.

For negative returns, a negative RDC can be calculated in the same way as for positive returns. In this case RDC quantifies the extent to which a negative return is so bad that it exhibits randomness deficiency in a downward direction. An investment with a high negative RDC can be shorted to produce profits, just in the same way that an investment with a positive RDC is longed. For example, the RDC of Facebook over the three months following its IPO on 18th May 2012 was -17, during which time it lost more than half of its value. The RDC scale therefore goes from -1 to $-\infty$ for investments generating negative returns and from $+1$ to $+\infty$ for investments generating positive returns. A risk-free investment has an RDC of 1.

We write *bootRDC* to specify when the bootstrapping definition is used to compute RDC. In the following section we establish the credibility of the bootstrapping method by proving that it converges with the PDF-based method for sufficiently large values of n, given that the returns are symmetrically distributed.

C. Proof that bootstrapping and PDF-based RDC converge

Consider the returns $R_1, R_2, \ldots, R_n$ of security t and suppose they are i.i.d. symmetrically distributed (i.e. n independant outcomes of random variable R, where R is centered at 0, note by substracting μ_t we can assume it is always the case). The above described bootstrapping method creates new i.i.d. random variables Y_i obtained by multiplying R_i by ± 1 depending on the outcome of $0-1$ valued coin flip c_i, i.e. $Y_i = (-1)^{c_i} R_i$.

The following result shows that the R_i's and Y_i's have the same distribution.

Lemma 1: If the returns $R_1, R_2, \ldots, R_n$ of security t follow a symmetric distribution with continuous density function (centered at 0) then the e.c.d. of the "coin flipped returns" $Y_1, Y_2, \ldots, Y_n$ a.s. converges to the e.c.d. of the returns.

Proof: Let $R_1, R_2, \ldots, R_n$ be the returns of security t as described above, centered at 0, Given n observed values for R_i's $r_1, r_2, \ldots, r_n$, the e.c.d of R is

$$F_n^R(x) = \frac{|\{i : \ r_i \leq x\}|}{n}$$

where x is any real number.

Let $\epsilon > 0$. Let n be large enough such that F_n^R is a good approximation of the c.d.f. F^R of R, i.e. by the Glivenko-Cantelli theorem (a.s. uniform convergence),

$$|F_n^R(x) - F^R(x)| < \epsilon/2 \quad \text{with probability 1.}$$

We claim that

$$|F_n^R(x) - F_n^Y(x)| < \epsilon \quad \text{with probability 1.}$$

Let us prove the claim. Let A be the measure 1 set of convergence given by the Glivenko-Cantelli theorem. Let x be in A and let $v_1, \ldots v_m$ be the subset of the $r_1, \ldots, r_n$ which are either less or equal to $-x$, or greater than x. Since F_n^R is close to F^R which is symmetrically centered at 0, the number of points occurring to the left of $-x$ and right of x is roughly the same, i.e. we have

$$\begin{aligned}
&|F_n^R(-x) - (1 - F_n^R(x))| \\
&\leq |F_n^R(-x) - F^R(-x)| + |F^R(-x) - (1 - F^R(x))| \\
&+ |(1 - F^R(x)) - (1 - F_n^R(x))| \\
&\leq \epsilon/2 + 0 + \epsilon/2 = \epsilon
\end{aligned}$$

i.e. there is at most $m/2 + \epsilon n/2$ v_i's to the left of $-x$.

The only y_i that can occur to the left of $-x$ are exactly the ones corresponding to the v_i's. Let $k = 2/\sqrt{\epsilon}$. By Chebychev inequality, given m coin flips (i.e. with mean $m/2$ and sd $\sigma = \sqrt{m}/2$), if $\bar{C}$ denotes the number of heads in m flips then

$$\Pr(|\bar{C} - m/2| > k\sqrt{m}/2) \leq \frac{1}{k^2}.$$

Thus with probability $1 - 1/k^2$ (over the coin flips) at least $m/2 - k\sqrt{m}/2$ y_i's land to the left of $-x$. Thus

$$\begin{aligned}
&|F_n^R(-x) - F_n^Y(-x)| \\
&\leq (1 - 1/k^2)|\frac{m/2 + \epsilon n/2}{n} - \frac{m/2 - k\sqrt{m}/2}{n}| + 1/k^2 \cdot 1 \\
&\leq \epsilon/2 + \epsilon/4 + \epsilon/4 = \epsilon
\end{aligned}$$

for n large enough, which proves the claim. Thus F_n^Y converges to F^R a.s. which proves the lemma. ■

By definition of RDC, the previous lemma shows that bootstrapping and PDF-based RDC agree for large values of n, i.e.

Theorem 1: Under $H_0^t : R_i \sim N(0, \sigma_t), \ i = 1, 2, \ldots, n$ for the returns $R_1, R_2, \ldots, R_n$ of security t, we have $bootRDC(t|H_0^t) = pdfRDC(t|H_0^t)$ for large enough n.

Proof: Given returns $R_1, R_2, \ldots, R_n$ of security t and H_0^t as above, let m_t^R be the sampled mean of returns computed from n observed values of R_i's. By Lemma 1, $\bar{Y}$ and $\bar{R}$ follow the same distribution under H_0^t, thus

$$\begin{aligned}
pdfRDC(t|H_0^t) &= \Pr(\bar{R} \geq m_t^R | \ H_0^t) \\
&= \Pr(\bar{Y} \geq m_t^R | \ H_0^t) = bootRDC(t|H_0^t)
\end{aligned}$$

■

D. Empirical Comparison of RDC Methods

We investigated the sample size n required to produce convergence between the two methods. The following R code was used for calculating PDF-based and bootstrapping RDC, where x is a sequence of historical values for a security.

```
RDC_pdfbased<-function(x){
   x<-diff(log10(x))
   prob<-pt((abs(mean(x))*sqrt(length(x)))/
   sd(x),df=length(x)-1)
   return(sign(mean(x))/(2*(1-prob)))
}
```

```
RDC_bootstrap<-function(x, precision=1000){
   x<-diff(log10(x))
   sums<-rowSums(matrix(  sample(c(-1,1),
   precision*length(x),replace=T),precision,
   length(x))*x)
   prob<-1-max(length(sums[sums>=abs(sum(x))])
   ,1)/length(sums)
   return(sign(mean(x))/(2*(1-prob)))
}
```

Figure 2 compares the two methods for increasing sample sizes n. Each point n on the x-axis shows the mean difference in ranks assigned by PDF-based and bootstrapping RDC for the S&P 500 stocks calculated using n daily returns starting January 1st 2011. The graph reveals that PDF-based RDC quickly converges on bootstrapping RDC, with no systematic divergence beyond $n \approx 20$. The correlation between the two sets of rankings is 0.997 when $n = 20$.

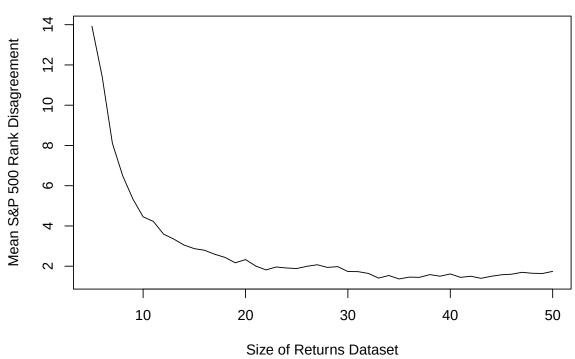

Fig. 2. Comparison of the two RDC methods for different historical dataset sizes

E. Applying RDC to historical data

Unlike other risk-to-reward measures, RDC allows the performance of risky investments to be easily compared against a risk-free investment. Table 2 shows the RDC for the U.S. stock market over the last 20 years based on the S&P total return index and using 13-week U.S Treasury Bonds as the risk-free rate. According to the capital asset pricing model (CAPM) model [9] [10] [11], investors can earn profits above the risk free rate by holding undiversifiable risk. Although stock market returns from the 1990s exhibited a level of high growth which, considered in isolation, might suggest an underlying performance above the risk-free rate, the returns during the 21st century (to July 2012) closely follow a random walk.

Period	Risk-Free Return	S&P 500 Return	RDC
1990 2000	59.8%	320%	118
2000 2010	30.1%	-8.2%	-1.62
1990 2010	108%	277%	3.14
2010	0.1%	13.2%	2.04
2011	0.1%	1.0%	1.03
21st century	30.4%	18.1%	-1.12

TABLE II
RDC VALUES FOR S&P 500 TOTAL RETURN INDEX

RDC can be used to analyse the extent to which any signal deviates from a random walk. For example, global average temperature from 1896 (the year the global warming hypothesis was originally proposed by Svante Arrhenius) to the present day yields an RDC of 131, using a 5-year rolling average to control for mean reversion. This high value indicates that the prediction of futher warming would make an outstanding investment.

F. Adjusted RDC

Assuming $\mu_B = R_f$, let us consider the case where an investor actively searches for the best investment by calculating the RDC of a variety of different securities. The security with the highest RDC is the one most likely to have an edge over the risk-free rate. It is therefore natural for investors to compare the RDCs of different securities and choose the best one. However, this selection process has the effect of altering the significance of the RDC by increasing the probability that a certain performance has been achieved through luck alone.

Whenever RDC is used to select one investment from among a group of investments, the RDC value must be adjusted. Imagine, for example, searching through all of the companies in the S&P 500 to find the one with the highest RDC. Based on 2011 returns, the stock with the highest RDC is Apple Inc. with a value of 242. This level of randomness deficiency is well beyond the level of significance typically required for scientific reporting (e.g. $p < .05$).

If Apple has been identified independently of the data used to calculate RDC, then this value would be very convincing. However, if Apple has been selected *because* of its performance (as opposed to some other independent hypothesis for why it should make a good investment), then we must reconsider the significance of its RDC. Specifically, we must evaluate the extent to which its RDC exceeds that which would be expected to arise by chance from within the group of securities being considered.

Given the hypothesis that all 500 S&P companies follow a random walk around the risk-free rate then, on average, half will exceed and half will underperform this benchmark. Since RDC quantifies the number of random permutations that would typically be required to match an observed performance, we would intuitively expect one of the S&P companies to have an absolute RDC of 250 by chance. In light of this expectation, Apple's RDC of 242 does not strongly refute the skeptical random walk hypothesis.

An advantage of RDC over other risk-to-reward measures is that it can be easily adjusted to account for this selection bias using the binomial distribution. Given the selection of a security t from among a pool of k candidates, then the adjusted RDC is given by

$$adjRDC(t) = \Pr(X \geq 1|H_0)^{-1} \quad \text{where}$$
$$H_0:\ X \sim Binom(p,k) \text{ where } p = RDC(t)^{-1}$$

This formula encapsulates the idea of choosing groups of random time series in batches of k and returns the number of batches that would, on average, have to be sampled before one is found to contain a random return matching the observed return. Table 3 shows the stocks with the highest RDCs based on returns over the last 10 years (2002 - 2012) alongside their adjusted RCDs. Returns have been adjusted for dividends, splits, mergers etc.

Company	Return	RDC	Adjusted-RDC
AAPL	3376%	139	1.20
CTSH	1910%	24.2	1.00
PCP	1116%	24.0	1.00
RRC	2071%	18.3	1.00
SWN	2435%	17.3	1.00

TABLE III
S&P 500 COMPANIES WITH HIGHEST RDCS (2002-2012) ADJUSTED FOR SELECTION BIAS

The data reveal that, when adjusted for selection bias, none of the stocks in the S&P 500 provide evidence of outperforming the risk-free rate over the last 10 years. Even the performance of Apple Inc., with a cumulative return of 3376% and an RDC of 139 is no better than would expect to find from looking through a set of 500 random walks. An investor who has identified Apple from among the S&P 500 companies because of its historical performance cannot expect it to continue to perform above the risk-free rate in the future, because there is no support for the hypothesis that its past performance was due to anything but luck.

If one is using historical 20th century returns from the U.S. stock market to justify the assumption that stock market returns reliably outperform the risk-free rate over the longer term (e.g. by pointing out that U.S. equities delivered an average of 4.3% inflation-adjusted annual return during this period), then the RDC values computed on these returns should also be adjusted for the number of national stock markets from which the U.S. was selected. For example, Jorion and Goetzmann [12] have argued that reliance on historical U.S. data for long-term estimates of expected returns is a serious problem. They argue that such estimates are subject to survivorship bias, in that the U.S. has been specifically identified by investors because of a historical level of economic growth which may not apply in the future. The high equity premium obtained for U.S. equities during the 20th century was the exception rather than the rule, with a much lower 0.8% return registered on average worldwide [12].

If, say, the RDC of the S&P 500 in the 1990s is adjusted for its selection from a pool of 56 national regulated exchanges affiliated with the World Federation of Exchanges then it falls from an original remarkable value of 118 to an adjusted value of only 4.7. It's possible that, in selecting the best-performing stock market in the world, investors who made significant profits from U.S. equities during the 20th century may have been relying more on luck than on prescience.

III. COMPARING RDC WITH SHARPE'S RATIO

Sharpe's ratio is currently the most common measure of risk-adjusted return and is widely used to rank the performance of portfolio and mutual fund managers [2]. Assuming that it works, the question arises as to whether it follows the same principles as RDC. In the following sections we prove that Sharpe's ratio converges with RDC under particular distribution-specific constraints.

A. Proof of Convergence of RDC and Sharpe's Ratio

Theorem 2: Let n be a fixed positive integer and consider m securities, with lognormally distributed returns over n days, with identical means. Then both RDC and Sharpe ratio will rank the securities in the same order.

Proof: Let u, t be two securities as above, with mean μ_m and sd σ_u and σ_t, such that $SR(u) \leq SR(t)$. We need to prove that $RDC(u|H_0^u) \leq RDC(t|H_0^t)$. By hypothesis we have

$$\frac{\log u_n/u_1}{\sigma_u} \leq \frac{\log t_n/t_1}{\sigma_t}$$

where u_1 and u_n (resp. t_1 and t_n) is the price of security u (resp. t) on day 1 and on day n. We have

$$\begin{aligned} RDC(u|H_0^u) &= \Pr(\bar{R}_u \geq m_u|H_0^u)^{-1} \\ &= \Pr(\bar{R}_u \geq \frac{\log u_n/u_1}{n}|H_0^u)^{-1} \\ &= \Pr(\frac{\sqrt{n}\bar{R}_u}{\sigma_u} \geq \frac{\log u_n/u_1}{\sigma_u\sqrt{n}}|H_0^u)^{-1} \\ &= \Pr(R \geq \frac{\log u_n/u_1}{\sigma_u\sqrt{n}}|H_0^u)^{-1} \end{aligned}$$

with $\bar{R}_u \sim N(\mu_m, \sigma_u/\sqrt{n})$ i.e., $R := \sqrt{n}\bar{R}_u/\sigma_u \sim N(\mu_m, 1)$. Thus

$$\begin{aligned} \Pr(R \geq \frac{\log u_n/u_1}{\sigma_u\sqrt{n}}|H_0^u)^{-1} &\leq \Pr(R \geq \frac{\log t_n/t_1}{\sigma_t\sqrt{n}}|H_0^t)^{-1} \\ &= \Pr(\frac{\sqrt{n}\bar{R}_t}{\sigma_t} \geq \frac{\log t_n/t_1}{\sigma_t\sqrt{n}}|H_0^t)^{-1} \end{aligned}$$

because $\sqrt{n}\bar{R}_t/\sigma_t \sim N(\mu_m, 1)$ since $\bar{R}_t \sim N(\mu_m, \sigma_t/\sqrt{n})$. Thus

$$\begin{aligned} \Pr(\frac{\sqrt{n}\bar{R}_t}{\sigma_t} \geq \frac{\log t_n/t_1}{\sigma_t\sqrt{n}}|H_0^t)^{-1} &= \Pr(\bar{R}_t \geq \frac{\log t_n/t_1}{n}|H_0^t)^{-1} \\ &= \Pr(\bar{R}_t \geq m_t|H_0^t)^{-1} \\ &= RDC(t|H_0^t) \end{aligned}$$

which proves the theorem. ■

The above proof demonstrates that RDC preserves Sharpe's ratio rankings as long as the returns are centered around the same value (e.g. the risk-free rate) and are normally distributed.

B. Time-frame Dependence

One limitation of the Sharpe ratio is that it is dependent on the time period over which it is measured [3]. Because rescaling modifies the standard deviation, the ratio can be increased by reporting returns over longer periods. To ensure a common standard it is common practice to annualize data that apply to periods other than one year, a process which is prone to error when returns are not normally-distributed.

A remarkable property of RDC is that, contrary to the Sharpe ratio, it is time-frame independent. The RDC of an investment calculated using a particular dataset remains consistent no matter what time periods are used to sample it (given the time period does not exceed the sample size).

Let t be a security whose logreturns $R_1, R_2, \ldots, R_n$ are normally distributed, and let k be an integer ($k << n$). We construct a new time series $Y_1, Y_2, \ldots, Y_{n/k}$ by taking the sum of groups of k values of R_i's, i.e. $Y_i = \sum_{j=(i-1)k+1}^{ik} R_j$. The time series Y is called a *k-regrouping of R*. The following result shows that RDC is preserved under k-regroupings.

Theorem 3: Let t be a security whose logreturns $R_1, R_2, \ldots, R_n$ are normally distributed, and let $k < n$ be an integer. Let $Y_1, Y_2, \ldots, Y_{n/k}$ be a k-regrouping of R. Computing the RDC of t under $H_0^R : R_i \sim N(\mu_t, \sigma_t),\ i = 1, 2, \ldots, n$ or $H_0^Y : Y_i \sim N(k\mu_t, \sqrt{k}\sigma_t),\ i = 1, 2, \ldots, n/k$ yields the same value.

Proof: Let $R_1, R_2, \ldots, R_n$, $Y_1, Y_2, \ldots, Y_{n/k}$, H_0^R and H_0^Y be as above. Given n observed values r_i for R_i's, the value of the sampled mean m_R is $m_R = \sum_i r_i$, and the value of the sampled mean m_Y satisfies $m_Y = km_R$. H_0^R implies $\bar{R} - \mu_t \sim N(0, \sigma_t/\sqrt{n})$ and H_0^Y implies $\frac{\bar{Y} - k\mu_t}{k} \sim N(0, \sigma_t/\sqrt{n})$. Thus,

$$
\begin{aligned}
RDC(t|H_0^R) &= \Pr(\bar{R} \geq m_R | H_0^R)^{-1} \\
&= \Pr(\bar{R} - \mu_t \geq m_R - \mu_t | H_0^R)^{-1} \\
&= \Pr(\frac{\bar{Y} - k\mu_t}{k} \geq m_R - \mu_t | H_0^Y)^{-1} \\
&= \Pr(\bar{Y} - k\mu_t \geq km_R - k\mu_t | H_0^Y)^{-1} \\
&= \Pr(\bar{Y} \geq km_R | H_0^Y)^{-1} \\
&= \Pr(\bar{Y} \geq m_Y | H_0^Y)^{-1} \\
&= RDC(t|H_0^Y)
\end{aligned}
$$

■

Since k-regrouping modifies the sd by a factor $\sqrt{k}$, Sharpe's ratio is not preserved under k-regrouping.

C. Ease of Interpretation

Cumulative return alone is not a sufficient statistic for identifying profitable investments, as this value fails to take into account the cost of the cash or loan which is needed to fund an investment in the first instance. As Sharpe points out [3], only a *differential* return information measure can be relied on to make the correct decisions.

Central to the usefulness of such measures is the fact that a differential return can be framed as a zero-investment strategy, that is, one which involves a zero outlay of money in the present and some potential profit or loss in the future, depending on circumstances. For example, an investor can take a long position in one asset and a short position in another (i.e. the benchmark), with the funds from the short position used to open the long position. If the latter provides a genuine differential return then profit can be derived with zero investment. RDC and Sharpe's ratio are both differential return information measures, although they differ in their ease of interpretation.

Table 4 gives the RDCs and annualized Sharpe ratios calculated for a selection of large cap stocks based on returns over the last 10 years (2002 - 2012).

Company	Return	Annualized SR	RDC
MCD	382%	0.61	13.7
KO	91.2%	0.23	2.1
XOM	167%	0.31	2.9
WMT	20.3%	0.00	1.0
AAPL	3376%	1.01	139

TABLE IV
COMPARISON OF RDC AND ANNUALIZED SR FOR PERIOD 2002 - 2012

As can be seen, both RDC and Sharpe's ratio identify that Walmart only manages to match the risk-free rate over the period, with both measures returning their lowest possible value for a profitable investment (1 and 0 respectively). As we proved earlier, the RDC and Sharpe ratio rankings are in agreement, yet the significance of the latter's values are less clear.

Apple Inc.'s RDC of 139 means that, on average, one would have to look through 139 profitable securities following a random walk about the risk-free rate before finding one with a performance this good. The Sharpe ratio of 1.01 means that the stock's annualized standard deviation closely matches its annualized excess return. To properly appreciate the significance of the latter measure an investor would need to hold extensive expertise on the relationship between mean displacement and standard deviation for the normal distribution, with regard to annualized values. RDC bypasses the need for such esoteric statistical knowledge by presenting the same information in terms of probability, which is more intuitive for investors.

Apple's Sharpe ratio is only 66% greater than that of McDonald's, which might suggest that it is perhaps twice as good an investment. However, Apple's RDC is 10 times greater than that of McDonald's, which means that it is 10 times more likely to be exhibiting underlying performance above the risk-free rate. The stock's overall return is also 9 times greater than that of the McDonald's Corporation. This substantial difference is not well communicated by Sharpe's ratio.

IV. CONCLUSION

Due to its many limitations, there are numerous strategies that fund managers can use to artificially enhance their Sharpe ratio and mislead investors. Since the ex post is often calculated using relatively few data points, fund managers can exploit strategies with skewed or kurtotic return distributions to distort the variance in their favour.

When financial variables are significantly non-normal, it becomes challenging to accurately compute the projected annualized skewness and kurtosis which Sharpe's ratio requires. Fama [13] originally analysed the distribution of the stocks making up the Dow Jones Industrial Average and concluded that security returns are 'fat tailed' (i.e. have kurtosis). More recent analyses of the S&P 500, FTSE 100, DAX and NIKKEI 225 for the period 2000 to 2009 have revealed extensive evidence of kurtosis, rejecting the null hypothesis of a normal distribution [14]. Analyses of hedge fund returns have also demonstrated negative skewness and positive excess kurtosis [7].

The time-frame dependence of Sharpe's ratio also allows the measure of risk to return to be misconstrued via the selective lengthening of the interval used to measure standard deviation or by smoothing monthly gains and losses by using derivative structures such as average price options, thereby reducing reported volatility [8].

RDC holds several advantages over Sharpe's ratio. First, and perhaps most importantly, where a sparse dataset renders it impossible to reliably identify the parameters of an underlying distribution (e.g. mean, standard deviation) the bootstrapping method, which does not rely on a probability distribution function, can be used instead. Because it is time-frame independent, RDC cannot be manipulated by computing standard deviation over different periods, and it also avoids the problem of having to annualize non-normal returns. Perhaps the most significant advantage of RDC over Sharpe's ratio is that the significance of its value is more intuitive and, as a result, can be more easily adjusted for selection bias.

In conclusion, we have described a risk-to-reward measure which is derived from the underlying relationship between risk and reward. We have proved that this measure converges with the Sharpe ratio under particular constraints, while holding the added advantages of being distribution independent, time-frame independent, and applicable to small datasets. All investment strategies, no matter how mundane or exotic, are ultimately accountable to demonstrate that their returns are not subject to chance. As such, the maximization of RDC is a robust and universal objective by which the performance of any financial security can be meaningfully evaluated and contrasted.

REFERENCES

[1] W. Sharpe, "Mutual fund performance," *Journal of Business*, vol. January, pp. 119–138, 1966.

[2] F. Modigliani and L. Modigliani, "Risk-adjusted performance: How to measure it and why," *The Journal of Portfolio Management*, vol. 23, no. 2, pp. 45–54, 1997.

[3] W. Sharpe, "The Sharpe Ratio," *The Journal of Portfolio Management*, vol. 21, no. 1, pp. 49–58, 1994.

[4] K. Dowd, "Financial risk management," *Financial Analysts Journal*, vol. 55, no. 4, pp. 65–71, 1999.

[5] J. Treynor, "How to rate management investment funds," *Harvard Business Review*, vol. 43, pp. 63–75, 1966.

[6] M. Jensen, "The performance of mutual funds in the period 1945-1964," *Journal of Finance*, vol. 23, pp. 389–416, 1968.

[7] D. Bailey and M. de Prado, "The Sharpe Ratio efficient frontier," *Journal of Risk*, vol. Forthcoming, 2012.

[8] R. Spurgin, "How to game your Sharpe Ratio," *The Journal of Alternative Investments*, vol. 4, pp. 38–46, 2001.

[9] H. Markowitz, "Portfolio selection," *The Journal of Finance*, vol. 7, no. 1, pp. 77–91, 1952.

[10] W. Sharpe, "Capital assets prices - a theory of market equilibrium under conditions of risk," *Journal of Finance*, vol. XIX, no. 3, pp. 425–442, 1964.

[11] J. Lintner, "The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets," *Review of Economics and Statistics*, vol. 47, no. 1, pp. 13–37, 1965.

[12] P. Jorion and W. Goetzmann, "Global stock markets in the twentieth century," *The Journal of Finance*, vol. LIV, no. 3, pp. 953–980, 1999.

[13] E. Fama, "The behaviour of stock prices," *Journal of Business*, vol. January, no. 37, pp. 934–105, 1965.

[14] M. Pesaran, "Predictability of asset returns and the efficient market hypothesis," *IZA Discussion Paper No. 5037, Insitute for the Study of Labor*, 2010.