2012

A Strategy-Proof Test of Portfolio Returns

Dean P. Foster
University of Pennsylvania

H. Peyton Young

A Strategy-Proof Test of Portfolio Returns

Abstract

Traditional methods for analyzing portfolio returns often rely on multifactor risk assessment, and tests of significance are typically based on variants of the t-test. This approach has serious limitations when analyzing the returns from dynamically traded portfolios that include derivative positions, because standard tests of significance can be 'gamed' using options trading strategies. To deal with this problem we propose a test that assumes nothing about the structure of returns except that they form a martingale difference. Although the test is conservative and corrects for unrealized tail risk, the loss in power is small at high levels of significance.

Disciplines
Statistics and Probability

A Strategy-Proof Test of Portfolio Returns

Dean P. Foster and H. Peyton Young

July 27, 2011

Abstract. Traditional methods for analyzing portfolio returns often rely on multifactor risk assessment, and tests of significance are typically based on variants of the t-test. This approach has serious limitations when analyzing the returns from dynamically traded portfolios that include derivative positions, because standard tests of significance can be 'gamed' using options trading strategies. To deal with this problem we propose a test that assumes nothing about the structure of returns except that they form a martingale difference. Although the test is conservative and corrects for unrealized tail risk, the loss in power is small at high levels of significance.

1. Gaming portfolio returns

A fundamental problem for investors is to determine whether a given portfolio is making positive returns relative to a benchmark such as the risk-free rate or a stock market index. This is a challenging problem for traditional statistical tests when little or nothing is known about the composition of the portfolio, the trading strategies the manager is using, or the amount of leverage he is taking on. The difficulty is that the manager may have both the incentive and ability to create returns that 'look good' for extended periods of time even though *in expectation* they are no better than the returns obtainable from standard market instruments. In particular, the manager can inflate his returns using options trading strategies that hide large downside risks in the tail of the distribution [Lo, 2001; Foster and Young, 2010]. Standard measures of performance, such as Jensen's alpha and the Sharpe ratio, do not take account of this unobserved downside risk. Nor do more recent proposals, such as the class of performance measures suggested by Goetzmann, Ingersoll, Spiegel, and Welch (2007).

To illustrate the difficulty, consider the returns series in Figure 1, which shows the monthly returns from a hypothetical portfolio over a period of 30 years. The returns appear to be i.i.d. normally distributed, and the OLS estimate of the mean monthly return over 360 months is 0.00324 (0.000764). The t-statistic is 4.2, which implies that the returns are positive with probability over 99.99%. In reality, however, the expected returns are zero. Here is how they were generated: in each month t the manager sold a covered

asset-or-nothing put[1] whose log probability of exercise, A_t, was determined by a random draw from a normal distribution with mean .00275 and standard deviation .0144. If the put was not exercised in a given month the manager earned an excess return of $e^{A_t} - 1$ over and above the risk-free rate. If the put was exercised in a given month, the fund would be completely wiped out. Assuming no arbitrage in the pricing of puts, the premium paid by the purchaser is offset by the probability that the put will be exercised and the expected excess return is zero. Under our assumptions the probability that the fund does not crash in a thirty-year period (360 months) is approximately $(e^{-.00275})^{360} \approx 0.37$.
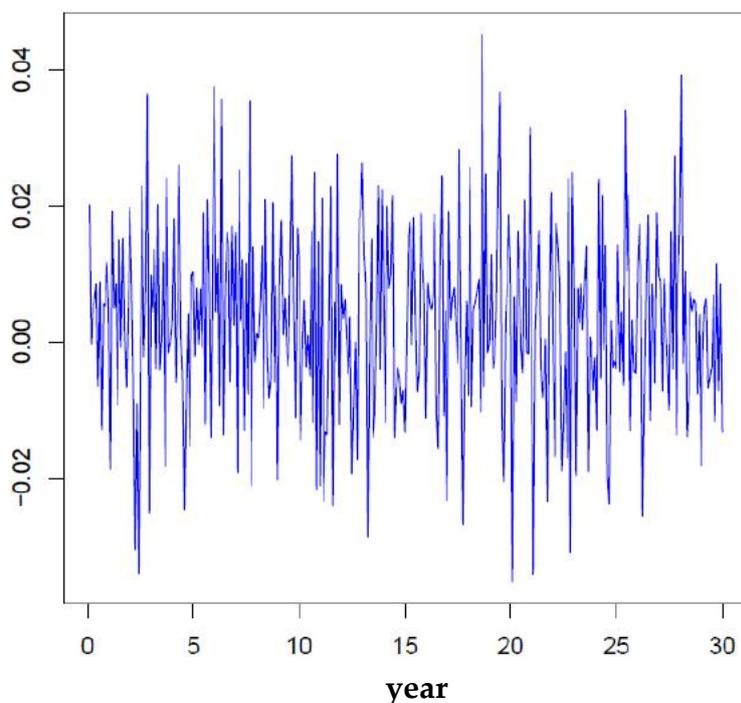

Figure 1. Monthly excess returns of a hypothetical portfolio over thirty years (360 months).

The particular simulation shown in Figure 1 is a series in which a crash did not occur. The *t*-test is fooled because the returns appear to be i.i.d. normal

---

[1] We assume that the put is covered by the portfolio itself, which will be completely liquidated if the put is exercised.

and positive on average, but there is a large potential loss that has not yet shown up in the data.

This type of manipulation is not purely hypothetical; indeed it is in the interest of fund managers to create return series that are "front-loaded" in order to attract clients and generate large performance bonuses (Lo, 2001; Foster and Young, 2010). The question is whether one can design a statistical test that protects against type-I errors, that is, against concluding that the returns exceed some given benchmark (such as the risk-free rate or a stock index) when in fact they do not.

The answer we propose is to test whether the *compound* excess returns form a nonnegative martingale. We first described this idea in its basic form in Foster, Stine, and Young (2008) and showed how to apply it to empirical returns data. The present paper is more theoretical and shows that we can greatly increase the statistical power of this class of tests through the use of leverage. The essential idea is the following: to test the returns from a given portfolio we construct a hypothetical family of leveraged versions of the portfolio. We then form a convex combination of these leveraged portfolios and apply the martingale maximal inequality to assess the probability that the returns from the original portfolio could have been produced by chance rather than superior expertise. This approach is similar in spirit to the universal portfolio framework pioneered by Cover (1991), but our construction is more specific and allows us to calculate the power of the test explicitly. The test is "strategy-proof" in the sense that a manager whose strategy does not produce excess returns will fail the test with high probability. (We give a formal definition of this concept in the next section.) However, the test is asymptotically powerful in the sense that, for small p-values, it will pass a series of bona fide excess returns with a probability that

is nearly as high as the *t*-test.

2. The Compound Excess Returns Test

Consider a portfolio whose returns $A_1, A_2, ..., A_T$ are observed at regular intervals (e.g., quarterly, monthly, or daily). We assume that these returns accurately reflect the portfolio's change in value in each period, but that the process producing the returns is a 'black box', that is, the manager's strategy is unobservable. If r_{ft} is the risk-free rate in period *t*, the *total excess return* in the period is $(1 + A_t)/(1 + r_{ft})$ and the *net excess return* is $\tilde{A}_t = (1 + A_t)/(1 + r_{ft}) - 1$.[2] We wish to test whether these returns are likely to have been produced by a strategy that yields positive excess returns in expectation, or whether they could have been produced by a strategy designed to fool us into thinking they are positive in expectation when in fact they are not.

The null hypothesis is that the excess returns have zero conditional expectation in every period, that is,

Null hypothesis
$$E[\tilde{A}_t \mid \tilde{A}_1, \tilde{A}_2, ..., \tilde{A}_{t-1}] = 0 \ for \ all \ t, \ 1 \le t \le T . \tag{1}$$

The alternative is that the expected net excess returns are nonnegative and in some periods they are strictly positive, that is,

Alternative hypothesis

$$E[\tilde{A}_t \mid \tilde{A}_1, \tilde{A}_2, ..., \tilde{A}_{t-1}] \ge 0 \ and \ strict \ inequality \ holds \ for \ some \ t, \ 1 \le t \le T. \tag{2}$$

---

[2] One can also define the excess return relative to other benchmarks, such as the return from a broad-based stock market index.

Conservative test. A hypothesis test is *conservative at significance level p* if it rejects the null with probability at most p when the null is true. A test is *conservative* if this property holds for all $p \in (0,1)$.

Strategy proof test. A test for excess returns is *strategy-proof* if it is conservative for any returns process satisfying (1), that is, whenever the excess returns form a martingale difference.

We shall assume throughout that a portfolio cannot lose more than 100% of its value in any period, that is, $\tilde{A}_t \geq -1$ for every t.[3] Hence the null hypothesis can be reformulated as follows:

Null: $C_t = \prod_{1 \leq s \leq t} (1 + \tilde{A}_s)$ *is a nonnegative martingale with expected value* 1.   (3)

The *martingale maximal inequality* states that, for any nonnegative martingale with expectation 1, and for any time T and target value $x > 0$, the maximum of the values $C_1, C_2, ..., C_T$ is greater than x with probability less than $1/x$ (Doob, 1953). Therefore the following is a strategy-proof test for excess returns: given any time T and significance level $p \in (0,1)$, reject the null hypothesis at time T if and only if

$$\max_{1 \leq t \leq T} C_t > 1/p.$$ (4)

We shall call this the *Compound Excess Returns Test (CERT)*.

---

[3] If the portfolio manager has short positions he could lose everything and still owe money to his creditors. However, this is a risk borne by the creditors not the investors: we assume that an investor cannot lose more than 100% of the amount invested, which places a lower bound of -1 on the net return in each period.

Note that this test assumes nothing about the parametric distribution of returns within a period or the serial dependence of returns among periods. It is also extremely simple to compute. A particularly important feature of the test is that it corrects for unobserved tail risk. For example, consider the compound returns generated by the period-by-period returns in Figure 1. By the end of year 30 the maximum compound value that the portfolio ever achieved was 3.42 times the size of a fund compounding at the risk-free rate. Our test says that this will happen with probability $1/3.42 = 0.29$. Thus we cannot reject the null with reasonable confidence. (In contrast the t-test incorrectly rejected the null with very high confidence.)
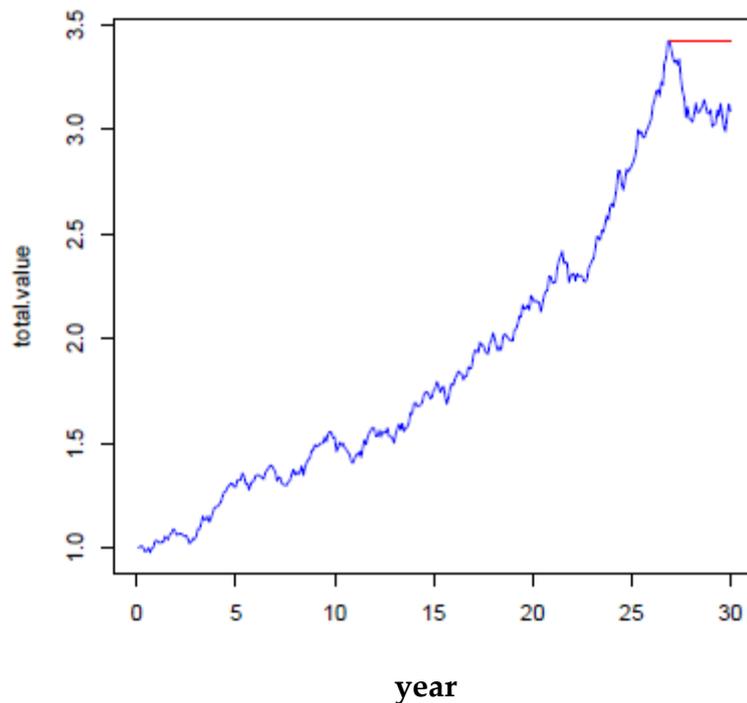

Figure 2. Compound value of the period-by-period returns shown in figure 1. The bar indicates the point at which the maximum value was achieved.

CERT is a highly conservative test but this is unavoidable if the test is to be strategy-proof. The reason is that any sequence of positive excess returns (of any length) can be reproduced by an options-trading strategy that has, in expectation, zero excess returns. The essence of the idea can be explained as follows (for further details see Foster and Young, 2010). Let $\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_T$ be a sequence of numbers that are the target excess returns over T periods. At the start of each period $t, 1 \leq t \leq T$, create a binary option that expires at the end of the period and gives the buyer of the option the right to the entire portfolio at the end of the period -- including the accumulated interest over the period at the risk-free rate plus the premium from the sale of the option itself. Design the option so that the probability of exercise is $p_t = \tilde{a}_t / (1 + \tilde{a}_t)$. Let v_t be the value of the portfolio at the start of the period (including the proceeds from selling the option), and let v_{t+1} be its value at the end of the period if the option is not exercised. This is a fair bet if

$$(1 - p_t)v_{t+1} + p_t \cdot 0 = (1 + r_{ft})v_t , \qquad (5)$$

that is,

$$v_{t+1} / v_t = (1 + \tilde{a}_t)(1 + r_{ft}) . \qquad (6)$$

This is a martingale strategy that produces the excess return $\tilde{a}_t$ with probability $1/(1 + \tilde{a}_t)$ and a total loss with probability $\tilde{a}_t / (1 + \tilde{a}_t)$. Repeated over T periods it produces the target sequence of returns $\tilde{a}_1, \tilde{a}_2, ..., \tilde{a}_T$ with probability $[\prod_{1 \leq t \leq T} (1 + \tilde{a}_t)]^{-1}$. To guard against such strategies – and a great variety of similar ones – one needs a test that is conservative with respect to the entire class of nonnegative martingales. *CERT* is such a test.

The preceding construction shows that the *CERT* rejection threshold is sharp:

8

a 'gamester' can produce a series of excess returns whose compound value grows by the a factor *1/p* with probability *p*. This has some striking implications. It says, for example, that to be 95% confident that a given 'black box' portfolio is producing positive excess returns, its compound value must grow by a factor of at least *twenty-fold* compared to a portfolio compounding at the risk-free rate. Although this may seem like an impossibly high standard to meet, it can be interpreted instead as the handicap that accompanies a complete lack of transparency. If the portfolio manager were to reveal more information about his positions and their implied tail risk, he might not need to meet such a high standard in order to convince investors that his returns are 'for real.' In subsequent sections we shall explore the effect of greater transparency in more detail, and show that the power of the test can be greatly increased if we have some information about extreme tail risk.

3. Related literature

There is a related literature on the testing of experts that spans both game theory and finance (Foster and Vohra, 1998; Lehrer, 2001; Shafer and Vovk, 2001; Sandroni, 2003; Sandroni, Smorodinsky and Vohra, 2003; Olszewski and Sandroni, 2008, 2010). This literature emphasizes the difficulty of identifying phony experts who make probabilistic forecasts of future events without having any knowledge of the process that is actually governing these events. There are many ways of formalizing this problem and we shall not attempt to review them here. Broadly speaking, however, our approach differs from this literature in two key respects. First, we restrict ourselves to a particular class of stochastic processes (non-negative martingales representing returns from a financial asset) and ask whether a supposed expert is producing returns from this class or not. Second, we adopt a classical hypothesis-testing approach to

assess how likely it is that a given series of returns was produced by an expert (at a given level of significance). In particular, this puts the burden of proof on the experts, who must distinguish themselves from the non-experts by producing returns that are highly unlikely to have been produced by chance. In much of the expert testing literature, by contrast, the null hypothesis is effectively reversed, and the expert is presumed to be expert unless the evidence is strongly against it. We would argue that, in the context of financial markets, it is fundamentally very difficult to consistently deliver excess returns because arbitrage opportunities tend to be eliminated through competition. Hence the natural presumption is that any given portfolio manager does not have such expertise until proven otherwise.

There is also a literature in finance that draws attention to the manipulability of traditional measures of performance, such as the Sharpe ratio and Jensen's alpha (Goetzmann, Ingersoll, Spiegel and Welsh, 2007). These authors propose a novel class of measures that overcome some forms of manipulation. This class is defined as follows: given a series of period-by-period returns $(\tilde{a}_1,...,\tilde{a}_T)$ and a parameter $\rho > 1$, define the function

$$G(\tilde{a}_1,...,\tilde{a}_T) = (1-\rho)^{-1} \ln[(1/T) \sum_{1 \le t \le T} (1+\tilde{a}_t)^{1-\rho}]. \qquad (7)$$

Since $1-\rho$ is negative, this measure imposes a heavy penalty on realizations in which any of the numbers $1+\tilde{a}_t$ is close to zero. The measure has the desirable property that it is invariant to the shifting of returns between periods, i.e., it makes no difference whether high returns come early or late in the sequence. However, it is quite different from our approach because it does not provide a test of significance – a metric for evaluating how *likely* it is that a given series of returns could have been produced by chance. Furthermore, unlike our test, it does not correct for unobserved tail risk.

The idea of using the martingale maximal inequality to test returns series was proposed in Foster, Stine and Young (2008). This paper shows how the martingale approach can be applied to the analysis of actual returns series from mutual funds and stocks. The present paper takes the analysis considerably further by showing how to ramp up the power of the approach by testing a convex combination of leveraged portfolios. The present paper is also related to an earlier one on the gaming of performance fees (Foster and Young, 2010). In that paper we showed that it is essentially impossible to design bonus schemes that reward expert managers who produce excess returns, and do not reward non-experts who cannot produce such returns. In other words, there are no *monetary* incentive schemes that induce managers to self-select into the expert and non-expert types. By contrast, the present paper shows that there exist *statistical* criteria of performance that can in fact distinguish between the two types.

4. Leveraging the compound excess returns test

The compound excess returns test guards against several different types of manipulation. The one that we have emphasized so far is unrealized tail risk, that is, the possibility that the fund could suddenly go bankrupt due to a low probability event that is hidden from investors. However, even if extreme tail risk is negligible there are other reasons why portfolio returns may be highly erratic and difficult to evaluate using standard statistical methods. Some of these problems arise from common trading strategies, such as market timing or momentum based strategies. Others may result from more deliberate manipulation, such as engaging in high leverage early in the returns series and reducing the leverage later on in order to enhance a performance measure such as the Sharpe ratio.

In this section we shall propose a variant of *CERT* that can be applied when extreme tail risk is not an issue, either because the manager offers a guarantee against large losses, or because the investor can control the downside by purchasing portfolio insurance. This still leaves the possibility that the returns are highly nonstationary and manipulated in other ways. We shall demonstrate a variant of our test that can be applied in this case that is *asymptotically as powerful as the t-test,* yet makes no assumptions about the serial correlation of returns or the parametric form of the returns-generating process.

The key idea is the following: when extreme tail risk is either absent or can be controlled through the purchase of portfolio insurance, the portfolio can be leveraged. We claim that the ability to leverage greatly increases the ability of the test to discriminate between processes that generate positive excess returns and those that do not. Furthermore, the level of leverage that optimizes the power of the test does not need to be known in advance: one can be completely ignorant about the optimal amount of leverage to use and still design a test that is very powerful at high levels of statistical significance. Like the unleveraged test (*CERT*) this approach guards against type-I errors, i.e., falsely concluding that a returns series is produced by an 'expert' manager when in fact it is produced by chance. It improves on *CERT* in the sense that type-II errors have low probability when the returns are well-behaved, i.e., satisfy the usual assumptions of independence and log normality.

As before, let $\tilde{A}_1, \tilde{A}_2, ..., \tilde{A}_t$ be the excess returns from a portfolio in each of t periods, net of the risk-free rate. We shall assume that the portfolio can be

insured against large downside losses over short periods. This can be done in several different ways. One possibility is for the portfolio manager to purchase insurance under a customized policy with an insurer. To obtain such a policy the manager would normally have to provide the insurer with considerable transparency regarding his positions and trading strategies. An alternative is for the investor to purchase protection against downside risk in the derivatives market (assuming such derivatives are available). This also presumes a certain amount of transparency, for otherwise market participants would have no way of knowing how to price the tail risk. In what follows we shall adopt the second point of view. While this is a somewhat restrictive assumption, it allows us to separate the issue of unrealized tail risk from the extent to which returns may be manipulated in other respects.

Let Δ be the length of a reporting period, e.g., a quarter, a month, or a day. Assume for simplicity that the price of the portfolio at the start of a given period t is \$1 per share. Given a number $0 \leq b < 1$, let $\gamma_t(b, \Delta)$ be the cost of a European put with strike price $b(1 + r_{ft})$ that expires at the end of the period. Suppose that we buy enough puts so that $1/(1 + \gamma_t(b, \Delta))$ of the portfolio is in shares and $\gamma_t(b, \Delta)/(1 + \gamma_t(b, \Delta))$ is in puts. Then all shares are protected and the value of the portfolio at the end of the period will be at least $b(1 + r_{ft})/(1 + \gamma_t(b, \Delta))$ times its value at the start of the period.

Given $\lambda > 1$, we can leverage this insured portfolio as follows: buy λ dollars of the insured portfolio and borrow $1 - \lambda$ dollars at the risk-free rate. Per dollar in the portfolio at the start of the period we will then have at least the following amount by the end of the period

$$\lambda b(1 + r_{ft})/(1 + \gamma_t(b, \Delta)) + (1 - \lambda)(1 + r_{ft}). \qquad (8)$$

For this to be nonnegative, it suffices that

$$b / (1 + \gamma_t(b, \Delta)) \geq 1 - 1/\lambda . \qquad (9)$$

Define $b_t(\lambda)$ to be the value of b that satisfies (9) as an equality. When the cost of puts is very small, which will be the relevant case in the results to follow,

$$b_t(\lambda) \approx 1 - 1/\lambda . \qquad (10)$$

For each $\lambda > 1$, let the random variable $B_t^{\Delta}(\lambda)$ denote the return (net of the risk-free rate) from the insured, leveraged portfolio constructed as above. Note that $B_t^{\Delta}(\lambda)$ is obtained by truncating the total return $1 + A_t$ below $b_t(\lambda)(1 + r_{ft})$, correcting for the cost of the puts, and dividing by $(1 + r_{ft})$. This leads to the expression

$$B_t^{\Delta}(\lambda) = \frac{\max\{b_t(\lambda), 1 + \tilde{A}_t\}}{1 + \gamma_t(b_t(\lambda), \Delta)} - 1 . \qquad (11)$$

By construction we know that

$$1 + \lambda B_t^{\Delta}(\lambda) \geq 0 . \qquad (12)$$

We shall say that options are *competitively priced* if the options market is efficient and there are no arbitrage opportunities. (This is their risk-neutral valuation.) In this case we must have

$$E[\tilde{A}_t] = E[B_t^{\Delta}(\lambda)], \qquad (13)$$

for otherwise one could make excess returns by selling portfolio insurance. It follows that, in a competitively priced options market,

$$\prod_{1 \le s \le t} (1 + \tilde{A}_s) \text{ is a nonnegative martingale iff } \prod_{1 \le s \le t} (1 + \lambda B_s^\Delta(\lambda)) \text{ is a}$$

$$\text{nonnegative martingale.} \tag{14}$$

In practice, of course, we cannot assume that options will be priced at exactly their competitive (risk-neutral) value. Fortunately we will not need to assume this for most of our subsequent results to hold: indeed our results on power hold even if out-of-the-money options are overpriced by a very large factor (see theorem 2 below).

There is, however, very little reason to think that options are *under-priced*, that is, portfolio insurance is too cheap. In other words, it is reasonable to suppose that in general

$$E[\tilde{A}_t] \ge E[B_t^\Delta(\lambda)]. \tag{15}$$

This condition implies that $\prod_{1 \le s \le t} (1 + \lambda B_s^\Delta(\lambda))$ is a *nonnegative supermartingale*. We shall assume that this condition holds for the remainder of the paper.

Theorem 1. *Let $\tilde{A}_1, \tilde{A}_2, ..., \tilde{A}_T$ be the excess returns of a portfolio over T periods of length $\Delta = 1/T$. Let $B_t^\Delta(\lambda)$ be the excess return in period t from an insured version of the portfolio that is leveraged at level $\lambda > 0$, and let $G(\lambda)$ be any distribution function for λ. The null hypothesis is that the returns $\tilde{A}_t$ are zero in expectation. The null can be rejected at significance level p if*

$$\max_{1 \le t \le T} \int_0^\infty [\prod_{1 \le s \le t} (1 + \lambda B_s^\Delta(\lambda))] dG(\lambda) > 1/p. \tag{16}$$

15

Proof. For each $\lambda > 0$ let

$$C_t^{\Delta}(\lambda) = \prod_{1 \leq s \leq t} (1 + \lambda B_s^{\Delta}(\lambda)). \qquad (17)$$

The null hypothesis is that $\prod_{1 \leq s \leq t} (1 + \tilde{A}_s)$ is a nonnegative martingale. Assuming that options are not underpriced, $C_t^{\Delta}(\lambda)$ constitutes a nonnegative supermartingale. Therefore $C_t^{\Delta} = \int_0^{\infty} \prod_{1 \leq s \leq t} (1 + \lambda B_s^{\Delta}(\lambda))] dG(\lambda)$ is a convex combination of nonnegative supermartingales, which implies that it too is a supermartingale. Given any *p*-value, it follows from the martingale maximal inequality that $\max_{1 \leq t \leq T} C_t^{\Delta} > 1/p$ with probability less than p (Doob, 1953). This concludes the proof of theorem 1.

We shall call the test in (16) the *Leveraged Compound Excess Returns Test (LCERT) with distribution function $G(\lambda)$*.

5. Power and leverage

We shall now show that when the distribution function $G(\lambda)$ is judiciously chosen, *LCERT is nearly as powerful as the optimal test when the p-value is small.* Let $C_\tau(1)$ be the compound value of the original asset in continuous time τ starting from an initial value $C_0(1) = 1$ with λ = 1. Suppose that the asset is lognormally distributed, that is, for some μ and σ,

$$\log C_\tau(1) \text{ is } N((\mu - \sigma^2 / 2)\tau, \sigma^2 \tau). \qquad (18)$$

Here and in what follows we shall always assume that *returns are expressed net of the risk-free rate*. The lognormal distribution is consistent with the standard representation of asset returns as a geometric Brownian motion in continuous time, and is the basis for Black-Scholes options pricing (Campbell, Lo, and MacKinlay, 1997).

When the asset is leveraged by a constant factor $\lambda > 0$ in continuous time, its compound value at time τ, $C_\tau(\lambda)$, is lognormally distributed:

$$\log C_\tau(\lambda) \text{ is } N((\lambda\mu - \lambda^2\sigma^2/2)\tau, \lambda^2\sigma^2\tau). \tag{19}$$

Fix a target time τ^* at which to test the null hypothesis at significance level p. There is no loss of generality in choosing the time scale so that $\tau^* = 1$, which will be assumed throughout the remainder of the analysis. The optimal test of the null hypothesis ($\mu = 0$) versus the alternative ($\mu > 0$) is the one-sided t-test. The t-test rejects the null ($\mu = 0$) at time $\tau^* = 1$ if

$$Z_1 = \frac{\log C_1 + (\sigma^2/2)}{\sigma} > z_p \text{ where } N(z_p) = (1-p). \tag{20}$$

Notice that this test is *independent* of the amount of leverage λ. Hereafter we shall usually omit the subscript 1 on the variables Z and C, it being understood that these are the values at the time ($\tau^* = 1$) the test is conducted.

We are interested in the situation where the returns are lognormal but we do not know this a priori. (If we did we would use the t-test.) The question is how much power we lose by using *LCERT*, which is in principle more conservative than the t-test, because it makes no assumptions about the normality or independence of returns among periods.

Maximal power loss function. Given $\sigma > 0$, $\Delta > 0$, and $p \in (0,1)$, the *maximal power loss function* $L_\sigma^\Delta(p)$ is the maximum probability over all $\mu > 0$ that the *t*-test correctly rejects the null $(\mu = 0)$ at level p while *LCERT* incorrectly accepts the null at level p.

We shall show that there exist 'universal' distributions $G(\lambda)$ such that $L_\sigma^\Delta(p)$ is small when: i) the value of p is small; ii) the time increments Δ are short, iii) put options are competitively priced.

In fact we can weaken the last condition considerably. Say that puts are *conservatively priced* if there exists a constant $K \geq 1$ such that the cost of a put is at most K times its risk-neutral valuation.

Theorem 2. *Let $\tilde{A}_1, \tilde{A}_2, ..., \tilde{A}_T$ be the excess returns of a portfolio over T periods of length $\Delta = 1/T$. Let $B_t^\Delta(\lambda)$ be the excess return in period t from an insured version of the portfolio that is leveraged at level $\lambda > 0$. The null hypothesis is that the returns $\tilde{A}_t$ are zero in expectation. The null can be rejected at significance level p if*

$$\max_{1 \leq t \leq T} \{ \int_0^\infty [\prod_{1 \leq s \leq t} (1 + \lambda B_s^\Delta(\lambda)] \frac{d\lambda}{(1+\lambda)^2} \} > 1/p. \tag{21}$$

ii) Suppose in addition that portfolio insurance is conservatively priced and the returns $\tilde{A}_t$ come from a lognormal process. Given any $\varepsilon > 0$, if p is sufficiently small and Δ is sufficiently small given p, the maximal power loss from the test (21) is less than ε.

Before turning to the proof, several remarks are in order.

Remark 1. We know from theorem 1 that the first statement (21) holds for any cumulative distribution function $G(\lambda)$. The essential claim is that under the

18

particular c.d.f. $G(\lambda) = \lambda/(1+\lambda)$ plus fairly weak conditions on the cost of insurance, the maximal power loss goes to zero with p. Other distribution functions also have this property; the key is that the distribution have full support on the positive reals and be fairly 'flat.'

Remark 2. The test in theorem 2 is "universal" in the sense that it can be applied with no prior knowledge of the actual distribution of returns. In particular, there is no presumption that the distribution is lognormal. The theorem states that if the returns *happen to be* lognormal, we do not lose much power by applying our test. At the same time we protect ourselves against type-I errors in the event that the returns are *not* lognormal (this follows from theorem 1). This is a luxury that the *t*-test does not permit.

The universal aspect of the test is similar in spirit to Cover's pioneering work on universal portfolios (Cover, 1991). Cover showed that a convex combination of leveraged portfolios will grow at a rate that is asymptotically as fast as the growth rate of an optimally leveraged portfolio. This is true in our set-up as well. However our focus is on deriving an asymptotic bound on the maximum power loss relative to a lognormal distribution. This requires a more specific choice for the distribution of leverage levels; it also requires factoring in the cost of portfolio insurance, which is not a feature of Cover's framework.

Remark 3. The proof will show that it suffices to be able to insure the portfolio against large downside losses over short periods of time at a cost that is not *unboundedly larger* than the risk-neutral value of the options. In particular, we do not assume that the options market is so complete or well-priced so that one could simply *deduce* the distribution of returns from the options prices themselves.

Remark 4. The test is most powerful when the level of significance is high (the *p*-value is small). The logic of this may be explained as follows. Suppose for example that $p = .001$. Then the compound value of the leveraged portfolio must grow by a factor of 1000 to pass our test. This is clearly very demanding, but in this case the *t*-test is also very demanding. The substance of the argument is to show that when the returns are lognormally distributed and they pass the *t*-test at a high level of significance, then the leveraged compound value is likely to pass our test as well.

Moreover, we would argue that small *p*-values are relevant in the context of financial markets, where there are many funds to choose from. For example, if there are *N* funds and we want to know whether the best of them is able to beat the market, we must correct for the fact that *pN* of them will pass at significance level *p* purely by chance. The Bonferroni correction for multiplicity implies that to be 95% confident that the best of 100 funds is run by an expert who can beat the market, it must pass at a level of .0005.

Proof of theorem 2. The first statement of the theorem (21) follows immediately from theorem 1, expression (16). The essence of theorem 2 is the second statement, namely, that for this particular distribution of leverage levels the power loss is arbitrarily small when p and Δ are sufficiently small. This result will be established in two steps. First we shall show that the power loss would be small if the portfolio could be leveraged continuously ($\Delta = 0$); then we shall show that the conclusion still holds when Δ is *sufficiently* small but not zero. (Note that, in our framework, the portfolio cannot be leveraged continuously because time periods are discrete. If continuous leveraging were possible, then we would be assuming that the price of the portfolio can

be represented as a continuous-time process, which would restrict the distribution much more than we wish to do.)

For each $\lambda > 0$, let $C(\lambda) \equiv C_1(\lambda)$ denote the compound value of the portfolio at time $\tau*=1$ when it is continuously leveraged at level λ. By assumption,

$$\ln C(\lambda) = (\lambda\sigma)z + \lambda\mu - \lambda^2\sigma^2/2, \quad \text{where } z \text{ is } N(0,1). \tag{22}$$

Define the random variable

$$w = z + \mu/\sigma. \tag{23}$$

Completing the square, (22) can be rewritten as follows:

$$\ln C(\lambda) = w^2/2 - (\lambda\sigma - w)^2/2. \tag{24}$$

Let

$$C = \int_0^\infty C(\lambda)g(\lambda)d\lambda, \tag{25}$$

where

$$g(\lambda) = 1/(1+\lambda)^2. \tag{26}$$

From (24) and (25) we deduce that

$$C = e^{w^2/2} \int_0^\infty e^{-(\lambda\sigma - w)^2/2} g(\lambda)d\lambda. \tag{27}$$

Making the change of variable $z = \lambda\sigma - w$, we obtain

$$C = \frac{\sqrt{2\pi}}{\sigma} e^{w^2/2} \int_{-w}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma) dz. \tag{28}$$

Let $N(\cdot)$ denote the cumulative normal distribution. Then we can write

$$\int_{-w}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma) dz > \int_{-w}^{w} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma) dz$$

$$= E[g(z/\sigma + w/\sigma) \mid -w \leq z \leq w][N(w) - N(-w)]$$

$$> g(w/\sigma)[N(w) - N(-w)]. \tag{29}$$

The last inequality follows from the convexity of $g(\lambda)$ and Jensen's inequality. From (28)-(29) we obtain

$$C > \frac{\sqrt{2\pi}}{\sigma} e^{w^2/2} g(w/\sigma)[N(w) - N(-w)]. \tag{30}$$

Note that this derivation holds for *any* convex density $g(\cdot)$ on the positive reals. This fact allows our results to be extended to various other distributions, though for the sake of concreteness we shall conduct the remainder of the proof using the specific density $g(\lambda) = 1/(1+\lambda)^2$. In this case (30) takes the form

$$C > (\sqrt{2\pi}) e^{w^2/2} [\sigma/(\sigma+w)^2][N(w) - N(-w)]. \tag{31}$$

We now turn to the estimation of the maximal power loss function $L_\sigma^0(p)$ when $\Delta = 0$ (continuous leverage). Power loss occurs whenever *LCERT accepts* the null and the *t*-test *rejects*. On the one hand, *LCERT* accepts the null at significance level p if

$$\max\{C_\tau : \tau \le 1\} \le 1/p.$$
(32)

This obviously implies that $C \equiv C_1 \le 1/p$. On the other hand, the t-test rejects the null at level p if

$$w = z + \mu/\sigma > z_p, \text{ where } N(z_p) = 1 - p.$$
(33)

Therefore the following *overestimates* the probability of power loss, that is, $L_\sigma^0(p)$ *is less than*

$$P(z_p < w \text{ and } C \le 1/p).$$
(34)

From this and (30) we conclude that $L_\sigma^0(p)$ *is less than*

$$P(z_p < w \text{ and } (\sqrt{2\pi})e^{w^2/2}[\sigma/(\sigma+w)^2][N(w) - N(-w)] \le 1/p).$$
(35)

Consider the right-most inequality in (35), namely,

$$(\sqrt{2\pi})e^{w^2/2}[\sigma/(\sigma+w)^2][N(w) - N(-w)] \le 1/p.$$
(36)

Taking logs of both sides we can rewrite this as follows:

$$w^2 \le 2\ln(1/p) - \ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma}) - 2\ln(N(w) - N(-w)).$$
(37)

Define the number

$$c_p \equiv \sqrt{2\ln(1/p)} \,. \tag{38}$$

From (34)-(38) we deduce that *the maximum power loss at significance level p is strictly less than $P(w \in I_p)$, where the interval I_p is defined as follows*

$$I_p = \{w : z_p < w \le \sqrt{c_p^2 - \ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma}) - 2\ln(N(w) - N(-w))}\}. \tag{39}$$

Lemma 2.1. *The length of the interval I_p goes to zero as p goes to zero.*

Proof. Let $\varepsilon > 0$. First we shall show that

$$w \in I_p \Rightarrow w \le c_p + \varepsilon \text{ for all sufficiently small } p \,. \tag{40}$$

Then we shall show that

$$c_p - z_p < \varepsilon \text{ for all sufficiently small } p \,. \tag{41}$$

To establish (40), note that $w \in I_p$ implies $z_p < w$, which implies

$$N(w) - N(-w) > N(z_p) - N(-z_p) = 1 - 2p \,. \tag{42}$$

From this we conclude that

$$w^2 < c_p^2 - \ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma}) - 2\ln(1 - 2p) \,, \tag{43}$$

and hence

$$w - c_p < (w + c_p)^{-1}[-\ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma}) - 2\ln(1 - 2p)] \ . \tag{44}$$

We know that $z_p \to \infty$ as $p \to 0$. Since $w > z_p$, the right-hand side of (44) is smaller than ε for all sufficiently small p. This establishes (40).

To prove the lemma, it remains to establish (41), namely, $c_p - z_p < \varepsilon$ for all sufficiently small p. We can estimate the value of z_p using the tail approximation for the normal distribution [Feller, 1957, p.193]:

$$p = 1 - N(z_p) > \frac{e^{-z_p^2/2}}{\sqrt{2\pi}}(z_p^{-1} - z_p^{-3}) . \tag{45}$$

From (45) we deduce that

$$c_p^2 \equiv 2\ln(1/p) < z_p^2 + 2\ln(2\pi) + 2\ln(z_p^3) - 2\ln(z_p^2 - 1) . \tag{46}$$

Hence

$$c_p - z_p < [2\ln(2\pi) + 2\ln(z_p^3)]/(c_p + z_p) . \tag{47}$$

Clearly the right-hand side of (47) is less than ε when p is sufficiently small. This establishes (41). Together with (40), it follows that the length of the interval I_p is less than 2ε for all sufficiently small p. This concludes the proof of Lemma 2.1.

The preceding shows that, given any small $\varepsilon > 0$, for all sufficiently small p the power loss $L^0_\sigma(p)$ is less than

$$P(z_p < w \leq z_p + 2\varepsilon) . \tag{48}$$

By definition, $w = z + \mu/\sigma$ where z is $N(0,1)$, hence $L^0_\sigma(p)$ is less than

$$\max_\mu \int_{z_p - \mu/\sigma}^{z_p - \mu/\sigma + 2\varepsilon} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz . \tag{49}$$

When p is sufficiently small the right-hand side of (49) is less than $2\varepsilon/\sqrt{2\pi}$. We have therefore shown that, for each σ, the *maximum* power loss (over all μ) is arbitrarily small provided that p is sufficiently small and the portfolio is leveraged continuously.

It remains to be shown that this statement remains true when the portfolio is leveraged over discrete time intervals of sufficiently short duration. Recall that the number of discrete periods is $T = 1/\Delta$ where $\Delta > 0$ is the length of a period. Given leverage level $\lambda > 0$, the compound value of the leveraged insured portfolio at the time the test is conducted is

$$C^\Delta(\lambda) = \prod_{1 \leq t \leq T} (1 + \lambda B^\Delta_t(\lambda)) . \tag{50}$$

The overall value of the portfolio is

$$C^\Delta = \int_0^\infty \prod_{1 \leq t \leq T} (1 + \lambda B^\Delta_t(\lambda)) g(\lambda) d\lambda . \tag{51}$$

We wish to compare the following values at the time $\tau^* = 1$ when the test is conducted:

C = *the value of the continuously leveraged portfolio*

C^Δ = *the value of the discrete-time leveraged portfolio*

We claim that, when Δ is small, C^Δ / C is close to one with high probability in the region where power loss occurs.

To establish this claim, recall that we estimated power loss in the continuously leveraged case by writing (see expression (28))

$$C = \frac{\sqrt{2\pi}}{\sigma} e^{w^2/2} \int_{-w}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma)dz .$$ (52)

The integral can be broken into two parts as follows

$$\int_{-w}^{w} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma)dz + \int_{w}^{\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma)dz .$$ (53)

In our previous estimation of C we dropped the second term in this expression (see (29)). In particular, we showed that power loss is small even when we underestimate C by ignoring realizations of $z = \lambda\sigma - w$ that are greater than w. Let us define

$$\tilde{C} = \frac{\sqrt{2\pi}}{\sigma} e^{w^2/2} \int_{-w}^{w} \frac{e^{-z^2/2}}{\sqrt{2\pi}} g(z/\sigma + w/\sigma)dz .$$ (54)

In effect, $\tilde{C}$ results from truncating the distribution of λ's to those satisfying $z = \lambda\sigma - w \leq w$, that is, $\lambda\sigma \leq 2w$. From the earlier part of the proof we know that, when p is sufficiently small, $w \leq c_p + \varepsilon$, and in particular $w \leq c_p + 1$. Thus

$\tilde{C}$ = *the value of the continuously leveraged portfolio when the*
distribution of λ's is truncated at $\bar{\lambda} \equiv 2(c_p + 1)/\sigma$. $\qquad$ (55)

As before, let $\gamma(\lambda, \Delta)$ denote the cost of insuring one dollar's worth of the portfolio for a period of length Δ at a strike price that allows leveraging at level $\lambda > 1$. Since the process is i.i.d. we can assume this cost is the same in all periods. When $\gamma(\lambda, \Delta)$ is small, the required strike price is approximately equal to $1 - 1/\lambda$, as we observed earlier (see expression (10)).

Let $q(\lambda, \Delta)$ be the *probability* that a put option with these characteristics is exercised. In other words, $q(\lambda, \Delta)$ is the probability that the portfolio loses at least $1/\lambda$ of its value by the end of the period. By assumption, the logarithm of the portfolio returns over a period of length Δ have the distribution $N(\mu\Delta - \sigma^2\Delta/2, \sigma^2\Delta)$ for some $\mu \geq 0$. In an efficient market, the cost of such an option is the probability of exercise times the conditional value if exercised, under the risk neutral assumption that $\mu = 0$. Since the conditional value when exercised is less than 1, the *cost* $q(\lambda, \Delta)$ of such an option satisfies

$q(\lambda, \Delta) = P(x \leq \ln(1 - 1/\lambda))$ where $x = z\sigma\sqrt{\Delta} - \sigma^2\Delta/2$ and z is $N(0,1)$. $\qquad$ (56)

When $\lambda > 1$ we have the inequality $\ln(1 - 1/\lambda) < -1/\lambda$, hence

$$q(\lambda, \Delta) < P(x \leq -1/\lambda) = P(z\sigma\sqrt{\Delta} - \sigma^2\Delta/2 \leq -1/\lambda)$$
$$= P(z \leq -1/\lambda\sigma\sqrt{\Delta} + \sigma\sqrt{\Delta}/2)$$
$\qquad$ (57)

When Δ is small we have

$$-1/\lambda\sigma\sqrt{\Delta}+\sigma\sqrt{\Delta}/2 < -1/2\lambda\sigma\sqrt{\Delta}. \tag{58}$$

From (57), (58), and the standard approximation of the tail of the normal we conclude that for all sufficiently small Δ,

$$q(\lambda,\Delta) < 2\lambda\sigma\sqrt{\Delta}e^{-1/8\lambda^2\sigma^2\Delta}. \tag{59}$$

In fact, we are only assuming that options are *conservatively priced*, that is, for some constant $K\geq 1$, $\gamma(\lambda,\Delta)$ is at most K times the risk neutral price. Therefore $\gamma(\lambda,\Delta)$ is at most K times the probability of exercise $q(\lambda,\Delta)$, and hence

$$\gamma(\lambda,\Delta) \leq 2K\lambda\sigma\sqrt{\Delta}e^{-1/8\lambda^2\sigma^2\Delta}. \tag{60}$$

To leverage the insured portfolio at level λ requires buying λ puts per period for $T=1/\Delta$ periods. This reduces the compound growth of the portfolio by a factor of at most $1-(1-\lambda\gamma(\lambda,\Delta))^{1/\Delta}$, which by (60) is at most

$$2K\lambda^2\sigma\Delta^{-1/2}e^{-1/8\lambda^2\sigma^2\Delta}. \tag{61}$$

Note that when Δ is sufficiently small, this factor is smaller than ε for *all* $1 < \lambda \leq \bar{\lambda}$. Thus the cost of the options does not reduce the compound growth of the portfolio by very much when the time periods are short.

There are, however, two further "costs" to insuring the portfolio. First, there is a positive probability that the value of the insured portfolio will fall to zero at some time before the test is conducted; in fact this occurs in any period

29

where the options are exercised. (By contrast, the value of a continuously leveraged, lognormally distributed asset almost surely never reaches zero.) The second potential cost is that the value of the portfolio over each time period does not exactly follow a lognormal distribution with leverage level λ, because the quantities of the portfolio and the risk-free asset are not rebalanced to keep the leverage constant *within* each period. To complete the proof of theorem 2, we need to show that with high probability these factors have only a small effect on the cumulative value of the portfolio when the test is conducted.

Let us recall that

$\tilde{C}$ = *the value of the continuously leveraged portfolio when the*
distribution of λ 's is truncated at $\bar{\lambda} \equiv 2(c_p + 1)/\sigma$.

We wish to compare this with

$\tilde{C}^{\Delta}$ = *the value of the discrete-time leveraged portfolio when the*
distribution of λ 's is truncated at $\bar{\lambda} = 2(c_p + 1)/\sigma$.

The preceding argument shows that when Δ is sufficiently small, the cost of the puts reduces the value of the insured portfolio $\tilde{C}^{\Delta}$ by a factor less than ε (compared to $\tilde{C}$). This assumes however that the puts are not exercised: if in any period a put is exercised, the value of the insured portfolio drops to zero. From (59) we know that the probability of this event is at most

$$\max_{1 < \lambda \leq \bar{\lambda}}\{\Delta^{-1} q(\lambda, \Delta)\} < \max_{1 < \lambda \leq \bar{\lambda}}\{2\lambda\sigma\Delta^{-1/2}e^{-1/8\lambda^2\sigma^2\Delta}\}$$
$$\leq 2\bar{\lambda}\sigma\Delta^{-1/2}e^{-1/8\bar{\lambda}^2\sigma^2\Delta}$$

(62)

The latter can be made less than ε by taking Δ to be sufficiently small. We have therefore shown that for all sufficiently small Δ: i) the probability is less than ε that the discrete-time portfolio $\tilde{C}^\Delta$ will go bankrupt; ii) if it does not go bankrupt, the cost of the puts decreases the value by a factor less than ε compared to $\tilde{C}$.

Finally, we must consider the fact that the discrete-time portfolio does not grow at exactly the same rate as the continuously leveraged portfolio *within* each period. The difference between the two rates is small in any given period with high probability; we claim that the cumulative difference over $1/\Delta$ periods is also small with high probability.

To establish this claim, fix $\lambda \le \bar{\lambda}$. Let $R_t(\lambda)$ be the growth rate in period t of the portfolio when continuously leveraged at level λ throughout the period. Let $R_t^\Delta(\lambda)$ be the growth rate of the discrete-time portfolio when leveraged at level λ at the start of period t. The following shows that the cumulative difference between these two rates over T periods is very small with high probability.

Lemma 2.2. If Δ is sufficiently small, $\displaystyle\sum_{1 \le t \le T}(R_t^\Delta(\lambda) - R_t(\lambda)) < \varepsilon$ with probability at least $1 - \varepsilon$.

Proof. The growth rate of the continuously-leveraged portfolio in period t is

$$R_t(\lambda) = e^{\lambda Z_t \sigma \sqrt{\Delta} + \lambda \mu \Delta - \lambda^2 \sigma^2 \Delta / 2} - 1 . \tag{63}$$

The growth rate of the discrete-time portfolio in period t is

$$R_t^\Delta(\lambda) = \lambda(e^{Z_t \sigma\sqrt{\Delta} + \mu\Delta - \sigma^2\Delta/2} - 1).$$ (64)

Note that the random variable Z_t is $N(0,1)$ and is the same in the two expressions. From the Taylor's expansion $e^x = 1 + x + x^2/2 + \ldots$ we see that (63) can be written as follows

$$R_t(\lambda) = \lambda Z_t \sigma\sqrt{\Delta} + \lambda\mu\Delta - \lambda^2\sigma^2\Delta/2 + \lambda^2 Z_t^2 \sigma^2\Delta/2 + O(\Delta^{3/2})f(\mu).$$ (65)

The coefficient on the residual term, $f(\mu)$, depends on μ, λ, and σ, but the latter two are fixed whereas μ is not. Similarly we have

$$R_t^\Delta(\lambda) = \lambda Z_t \sigma\sqrt{\Delta} + \lambda\mu\Delta - \lambda\sigma^2\Delta/2 + \lambda Z_t^2 \sigma^2\Delta/2 + O(\Delta^{3/2})\tilde{f}(\mu).$$ (66)

Therefore we can express the difference between the growth rates as follows for some bounded function $h(\mu)$:

$$R_t^\Delta(\lambda) - R_t(\lambda) = (Z_t^2 - 1)(\lambda^2 - \lambda)\sigma^2\Delta/2 + O(\Delta^{3/2})h(\mu).$$ (67)

Next we estimate the cumulative difference over $T = 1/\Delta$ periods:

$$\sum_{1 \le t \le T}(R_t^\Delta(\lambda) - R_t(\lambda)) = .5(\lambda^2 - \lambda)\sigma^2\Delta \sum_{1 \le t \le T}(Z_t^2 - 1) + O(\sqrt{\Delta})h(\mu).$$ (68)

The T draws of Z_t are independent, $Z_t^2 - 1$ has mean zero and variance 4, hence

$$\sum_{1 \le t \le T}(Z_t^2 - 1) \text{ is approximately } N(0, 4/\Delta).$$ (69)

32

Therefore the overall difference in growth rates between the two portfolios is approximately

$$\sum_{1 \leq t \leq T} (R_t^\Delta(\lambda) - R_t(\lambda)) \approx W(\lambda^2 - \lambda)\sigma^2 \sqrt{\Delta} + O(\sqrt{\Delta})h(\mu) \text{ where } W \text{ is } N(0,1). \quad (70)$$

For a given value of μ, we can make this difference less than ε with probability at least $1 - \varepsilon$ by choosing Δ to be sufficiently small, say $0 < \Delta \leq \Delta_\mu$. This concludes the proof of Lemma 2.2.

To complete the proof of theorem 2, let us recall that we are estimating the loss of power in the worst case, that is, for values of μ that *maximize* the probability of accepting the null when the *t*-test rejects. When μ is sufficiently large, say greater than $\bar{\mu}$, the probability that our test accepts the null is less than ε, hence the power loss must also be less than ε. (This is an immediate consequence of expression (19).) Thus in expression (70) it suffices to consider values of μ satisfying $\mu \leq \bar{\mu}$. The function $h(\mu)$ is uniformly bounded above for all $\mu \leq \bar{\mu}$, hence there is a number $\bar{\Delta} > 0$ such that (70) holds uniformly for all Δ such that $0 < \Delta \leq \bar{\Delta}$.

Putting the various parts of the argument together, we have shown that there are numbers $\bar{\Delta} > 0$ and $\bar{\mu} > 0$ such that: i) power loss is less than ε when $\mu > \bar{\mu}$; ii) when $\mu \leq \bar{\mu}$ and $0 < \Delta \leq \bar{\Delta}$, the cumulative value of the insured portfolio C^Δ closely approximates the cumulative value of the continuously leveraged portfolio $\tilde{C}$ in the sense that

$$C^\Delta \geq (1 - 2\varepsilon)\tilde{C} \text{ with probability at least } 1 - 2\varepsilon. \quad (71)$$

We already know, however, that the maximum power loss goes to zero as p goes to zero for the continuously leveraged portfolio $\tilde{C}$. It follows from (71) that the same holds for the insured discrete-time portfolio C^{Δ}. This completes the proof of Theorem 2.

5. Targeting the leverage

Theorem 2 demonstrates the existence of a strategy-proof test that has good power even when one has no prior information about the variance of the underlying process. However, the maximal power loss function is difficult to calculate explicitly because of its dependence on the unknown σ and on the significance level p of the test. We can obtain a clearer picture of the magnitude of the power loss if we assume that σ is known with high accuracy to begin with. In this case we do not need to integrate over a distribution of leverage levels; we can instead choose a single level that optimizes the power of the test conditional on our estimate of σ. This situation is summarized in the following result.

Theorem 3. *Let* $\tilde{A}_1, \tilde{A}_2, ..., \tilde{A}_T$ *be the excess returns of a portfolio over T periods of length* $\Delta = 1/T$. *Let* σ^* *be an estimate of the standard deviation of the returns, and let p be the desired level of significance. Let* $B_t^{\Delta}(\lambda^*)$ *be the excess return in period t from an insured version of the portfolio that is leveraged at level*

$$\lambda^* = c_p / \sigma^* \ where \ c_p \equiv \sqrt{2\ln(1/p)} . \tag{72}$$

The null hypothesis is that the returns $\tilde{A}_t$ *are zero in expectation. The null can be rejected at level p if*

$$C^\Delta(\lambda^*) \equiv \prod_{1 \le t \le T} (1 + \lambda^* B_t^\Delta(\lambda^*)) > 1/p. \tag{73}$$

If portfolio insurance is conservatively priced and Δ is sufficiently small, and if the estimate σ^* is correct, the maximal power loss of this test is at most

$$2N(.5(z_p - c_p)) - 1. \tag{74}$$

Proof of theorem 3.

We shall estimate the power loss in the continuous case, from which the conclusion in the discrete case follows by arguments similar to those in the proof of theorem 2. When the portfolio is continuously leveraged at level $\lambda > 0$, its value at the time when the test is conducted ($\tau^* = 1$) is

$$\ln C(\lambda) = (\lambda\sigma)z + \lambda\mu - (\lambda\sigma)^2/2 \ \ where \ z \ is \ N(0,1). \tag{75}$$

Power loss occurs when $\ln C(\lambda) \le 1/p$ and the *t*-test rejects the null, that is, when

$$z_p < z \le \frac{\log(1/p) - \lambda\mu + (\lambda\sigma)^2/2}{\lambda\sigma}. \tag{76}$$

The probability of this event is minimized when the right-hand side is as small as possible, which occurs (for a given μ and σ) when $\lambda = c_p/\sigma$. Call this the *optimal amount of leverage*

$$\lambda^*(p,\sigma) \equiv c_p/\sigma. \tag{77}$$

35

Given our estimate σ^*, we will use the optimal amount of leverage provided our estimate of σ is accurate, which is assumed in the statement of the theorem. The important point is that the optimal leverage does *not* depend on μ (which is assumed to be unknown).

Substituting the optimal λ^* into (76) and rearranging terms, we conclude that power loss occurs when

$$z_p < z + \mu / \sigma \leq c_p. \tag{78}$$

For a given σ and p, the probability of event (78) is maximized when μ / σ lies halfway between z_p and c_p, that is, when

$$\mu = (z_p + c_p)\sigma / 2. \tag{79}$$

Therefore the power loss is *at most*

$$P(z_p < z + (z_p + c_p)/2 \leq c_p) = P(.5(z_p - c_p) < z \leq .5(c_p - z_p))$$

$$= N(.5(c_p - z_p)) - N(.5(z_p - c_p)) \tag{80}$$

$$= 2N(.5(z_p - c_p)) - 1.$$

However, we already know from (41) that $(z_p - c_p) \to 0 \ as \ p \to 0$. This completes the proof of theorem 3.

We are now in a position to compare the maximal power loss of our test when σ is known and when σ is unknown. The first case is covered by theorem 3 and the second by theorem 2. These are admittedly somewhat extreme

situations: in practice one might have a preliminary estimate (or guess) about σ that is subject to some uncertainty. In this case one could choose a universal distribution of leverage levels that is centered around one's best guess (as we shall show explicitly below), and one would be protected if one's guess turns out to be wrong.

To fix ideas let us begin by considering the maximal power loss when σ is known. In this case we have the upper bound

$$f(p) = 2N(.5(z_p - c_p)) - 1. \tag{81}$$

The solid curve in figure 3 shows the behaviour of $f(p)$ for small values of p. Of course this is only a theoretical upper bound on the maximal power loss, and it was obtained by making a number of approximations. The most significant of these was that we estimated the power loss using the compound value of the portfolio *at the time the test is conducted* rather than its *maximal value up to the time that the test is conducted*. The squares in figure 3 represent the power loss over many simulations when this weaker test is used. The close correspondence between the squares and the curve show that our estimate, $f(p)$, is quite close to the actual power loss using this weaker test.

When we simulate the process using the stronger test (based on the maximal compound value), we obtain the dotted line shown in Figure 3. This shows that the actual power loss from the test is considerably smaller than the theoretical upper bound in (81). For example, when p = .001 the maximal power loss (based on simulations) is on the order of 15%, whereas the upper bound we computed is about 25%. A 15% chance of a type-II error is a modest price to pay for the extra protection afforded by our test compared to

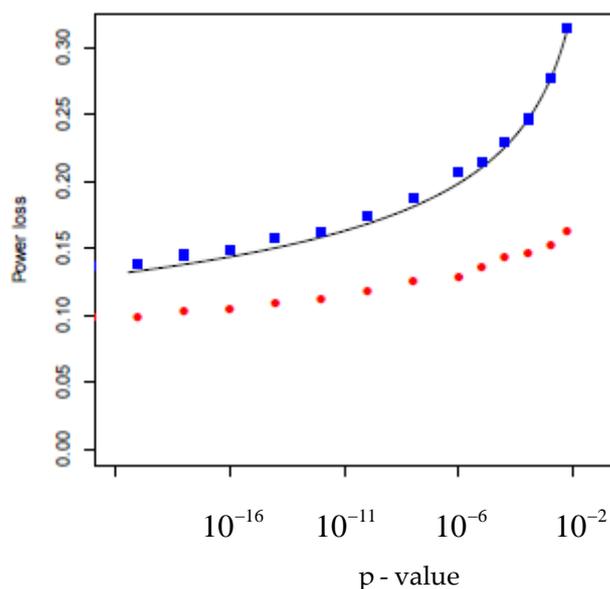the *t*-test, which, as we have seen, is easy to game.

Figure 3. Maximal power loss of *LCERT* as a function of the significance level *p*. The curve represents the theoretical upper bound. The squares represent the power loss as estimated from simulations when the weaker form of the test is used (the compound value of the portfolio at the time of the test). The dots represent the power loss as estimated from simulations when the stronger form of the test is used (the maximal compound value up to the time of the test).

We should stress that all of these results involve a worst-case analysis, that is, we are estimating the maximum power loss that could occur if the mean of the distribution happens to fall in a particular range. For most values of μ the power loss will be very small at a given level of significance. To illustrate this point, let us fix σ and p and compute the power loss for different values of μ. Since the loss depends only on the ratio μ/σ, we may as well assume that $\sigma = 1$. The results for p = .001 are illustrated in figure 4. Note that although the *maximal* power loss is about .15, it is less than .05 outside of the range $2 \le \mu \le 5$.
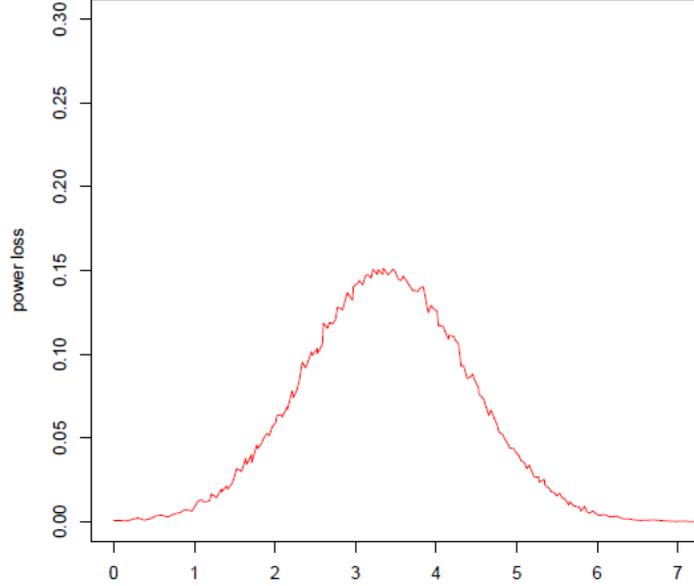
Figure 4. Power loss as a function of μ when $\sigma = 1$ and $p = .001$.

6. Comparing power loss under targeted versus distributed leverage

When we do not have a good prior estimate of σ we need to employ a universal test such as the one in theorem 2. In this case the power loss can be considerably higher. This is the price we must pay for lack of information about the variance of the true distribution. To estimate the magnitude of the difference, let us recall that under the universal test, power loss occurs in the interval

$$z_p < w \le \sqrt{c_p^2 - \ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma}) - 2\ln(N(w) - N(-w))}, \qquad (82)$$

where $w = z + \mu/\sigma$ and z is $N(0,1)$. Since $w > z_p$, the term $N(w) - N(-w)$ is at most 1 and at least $1 - 2p$. Hence when $p \le .01$ the quantity $2\ln(N(w) - N(-w))$ is negligible, and we have the nearly equivalent expression

$$z_p < w \le \sqrt{c_p^2 - \ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma})} \ . \tag{83}$$

This defines a power loss interval whose lower end-point is z_p and whose upper end-point is the value of w that solves

$$w^2 = c_p^2 - \ln(2\pi) + 4\ln(\sqrt{\sigma} + w/\sqrt{\sigma}) \ . \tag{84}$$

Call this value $d_p(\sigma)$. It is straightforward to show that $d_p(\sigma)$ is strictly larger than c_p for all σ. Indeed if we define $d_p = \min_\sigma d_p(\sigma)$, then d_p is the solution to the equation ,

$$d_p^2 = c_p^2 + \ln(d_p^2) + \ln(8/\pi) \ , \tag{85}$$

which is certainly larger than c_p .

Suppose now that we have an estimate or "guess" about the value of σ, say $\bar{\sigma}$, but we want to be protected against type-I errors if our guess turns out to be wrong. Let p be the desired level of significance. We can then rescale the universal distribution so that the smallest power loss interval occurs precisely when our guess is correct, that is, when $\sigma = \bar{\sigma}$. It can be shown that this is achieved by choosing the density

$$g(\lambda) = \frac{d_p/\bar{\sigma}}{((d_p/\bar{\sigma}) + \lambda)^2} \ . \tag{86}$$

It is straightforward to show that the minimum power loss interval (the

analog of (83)) occurs when $\sigma = \bar{\sigma}$, and this interval is precisely $(z_p, d_p]$.

We can now compute an upper bound on the power loss. This will depend, of course, on how far off our estimate $(\bar{\sigma})$ is from the truth (σ). There is no loss of generality in assuming that $\bar{\sigma} = 1$, in which case we can use the density $g(\lambda) = d_p / (d_p + \lambda)^2$. The preceding argument shows that the following is an upper bound on the maximum power loss for given p and σ:

$$\bar{L}_p(\sigma) \equiv 2N(.5(d_p(\sigma) - z_p)) - 1, \qquad (87)$$

where $d_p(\sigma)$ is the value of that solves equation (84).

Table 1 gives values of $\bar{L}_p(\sigma)$ for various choices of σ. Notice that they are very stable across radically different values of σ. For example, even if the estimate is off by a factor of 10, the power loss is only slightly higher than if the estimate is on target.

Let us compare these values with the power loss when σ is known in advance and we choose the single level of leverage (namely, $\lambda = c_p / \sigma$) that optimizes power given this σ. This is the solid curve in figure 3. For the range of p-values in table 1, the power increases by about .20. But what if our estimate of σ turns out to be wrong? In this case the loss of power can be much greater than if we used the universal distribution. This is shown in table 2 below, which gives the bound on power loss when our prior estimate is $\bar{\sigma} = 1$ and the true value is σ. Comparing these values with the corresponding values in table 1, we see that the universal distribution offers considerably more protection (lower power loss) when the prior estimate of σ is off by a factor of two or more. We should also remark that the numbers

41

in both tables are upper bounds on the maximal power loss function. As we saw in figure 3, the actual loss will be considerably less (on the order of .10 less) for this range of values of σ and p.

p	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
σ					
0.1	.57	.51	.47	.44	.42
0.2	.52	.47	.44	.41	.39
0.5	.48	.43	.40	.38	.36
1	.46	.42	.39	.37	.35
2	.47	.43	.40	.38	.36
10	.56	.51	.47	.44	.42

Table 1. Upper bound on the power loss when the true standard deviation is σ, the estimate is $\bar{\sigma} = 1$, and the distribution $g(\lambda) = d_p / (d_p + \lambda)^2$ is used.

p	10^{-2}	10^{-3}	10^{-4}	10^{-5}	10^{-6}
σ					
0.1	.99	1.00	1.00	1.00	1.00
0.2	.99	1.00	1.00	1.00	1.00
0.5	.54	.56	.59	.61	.64
1	.28	.25	.23	.21	.20
2	.54	.56	.59	.61	.64
10	1.00	1.00	1.00	1.00	1.00

Table 2. Upper bound on power loss when the true standard deviation is σ, the estimate is $\bar{\sigma} = 1$, and the single leverage level $\lambda = c_p$ is used.

7. Conclusion

In this paper we have proposed a test of portfolio returns that is appropriate when an investor has no knowledge of the portfolio manager's investment strategy. First, the test protects against the possibility that there is unrealized tail risk. This is important because managers typically have a strong incentive to take on such risk in order to enhance their performance bonuses. The level of risk can be low and still create the appearance of sizable excess returns before the risk is realized. For example, an annual excess return equal to 5% is impressive, yet it can be driven by tail risk that in expectation will take 20 years to show up. Second, the test is valid for returns that are serially dependent and not normally distributed, a situation that can easily result from dynamic trading strategies and market timing schemes. Third, while the test is inherently conservative, it can be turned into a powerful test if we have some information about the price of insuring against extreme tail risk. Contrary to what one might expect, the price of insurance need only be roughly correct -- say within an order of magnitude of the risk-neutral price -- for the test to have high power.

References

Campbell, John C., Andrew W. Lo, and A.C. MacKinlay, *The Econometrics of Financial Markets.* Princeton NJ: Princeton University Press, 1997.

Cover, Thomas M. (1991), "Universal portfolios," *Mathematical Finance* 1: 1-29.

Doob, Joseph L., *Stochastic Processes*, New York: John Wiley, 1953.

Feller, William, *An Introduction to Probability Theory and Its Applications,* Princeton University Press, 1971.

Foster, Dean P., Robert Stine, and H. Peyton Young (2008), "A Martingale Test for Alpha," Wharton Financial Institutions Center Working Paper 08-041.

Foster, Dean P., and H. Peyton Young (2010), "Gaming Performance Fees by Portfolio Managers*," Quarterly Journal of Economics*, 125: 1435-1458.

Goetzmann, William, Jonathan Ingersoll, Matthew Spiegel, and Ivo Welch (2007), "Portfolio Performance Manipulation and Manipulation-Proof Performance Measures," *Review of Financial Studies*, 20, 1503-1546.

Hull, John C., *Options, Futures and Other Derivatives, 7th Edition*. (New York: Prentice-Hall, 2009).

Lehrer, Ehud (2001), "Any Inspection Rule is Manipulable," *Econometrica*, 69, 1333-1347.

Lo, Andrew W. (2001), "Risk Management for Hedge Funds: Introduction and Overview," *Financial Analysts' Journal*, November/December 2001, 16-33.

Olszewski, Wojciech, and Alvaro Sandroni (2008), "Manipulability and Future-Independent Tests," *Econometrica*, 76, 1437-1480.

Olszewski, Wojciech, and Alvaro Sandroni, "A Non-Manipulable Test," *Annals of Statistics*, forthcoming.

Sandroni, Alvaro (2003), "The Reproducible Properties of Correct Forecasts," *International Journal of Game Theory*, 32, 151-159.

Sandroni, Alvaro, Rann Smorodinsky, and Rakesh Vohra (2003), "Calibration with Many Checking Rules," *Mathematics of Operations Research*, 28, 141-153.

Shafer, Glenn, and Vladimir Vovk (2001), *Probability and Finance: It's Only a Game*, New York: John Wiley.