

# Flaw in the fund skill/luck test method of Cuthbertson et al

Nuttall, John University of Western Ontario

22. January 2007

Online at http://mpra.ub.uni-muenchen.de/1584/ MPRA Paper No. 1584, posted 07. November 2007 / 01:51

# FLAW IN THE FUND SKILL/LUCK TEST METHOD OF CUTHBERTSON ET AL (SSRN Abstract 665744)

# John Nuttall

**University of Western Ontario** 

jnuttall@uwo.ca

January 20, 2007

## ABSTRACT

Cuthbertson et al have recently described a method that is claimed to be able to identify individual fund managers who exhibited skill over a long period in the past. The only input to the process is monthly fund returns. We suggest that a critical step in the Cuthbertson method is flawed. This step involves the study of the order statistics of period average fund returns. We construct a simple model to which the Cuthbertson method should apply. Simulations with the model conclusively demonstrate that the method fails to detect many funds with skill, and also erroneously identifies many funds as having skill they do not possess.

COMMENTS WELCOME

#### INTRODUCTION

In a recent working paper Cuthbertson et al [Cu05] have described a method (the CNO method) designed to identify, from a large group, a number of individual fund managers that exhibited skill over a long period in the past. The method uses only data on the monthly returns of a group of funds during the period.

In this note we describe a model containing a mixture of skilled and unskilled managers to which the essentials of the CNO method should apply. We perform a Monte Carlo simulation of the performance of the model. The calculations show that the CNO method gives highly misleading results for situations involving high performance stocks that might well be similar to those in real markets. We find that many managers identified as skilled by the method in fact have no skill, and many managers with skill are not identified.

We present an argument that explains why the CNO test for skill fails, based on the unsound nature of the null hypothesis at the heart of the method.

The techniques of CNO [Cu05] are in very large part the same as those of an earlier paper by Kosowski et al [Ko05], which we call KTWW. Apart from their application to different markets, from our point of view the main difference between the two articles is that CNO make a clear claim of being able to identify skill in individual managers, whereas KTWW restrict precise skill claims to an unknown subset in a specified set of funds. In relation to effects related to individual funds, KTWW do not make categorical claims about skill, but some of their statements could be taken to have this implication. In the section below headed Quotes we present some statements from the two papers so that readers may form their own opinion on this point. Apart from that, we concentrate in this note on the plausibility of the CNO claim only.

#### STRUCTURE OF THE KTWW/CNO METHOD

We study a collection of *n* funds. The actual performance of fund *k* over the period is represented by a specified quantity (e.g. alpha or its t-statistic) that we shall call return  $r_k, k = 1, ..., n$ . The reader may think of it as the average monthly return of the fund over the period.

By means of a bootstrap procedure, the details being of no concern for our present purpose, the two papers obtain a discrete distribution intended to approximate the probability distribution function (PDF) of the returns of the funds under the assumption that no fund employs skill. CNO call this the 'luck distribution', and display a version smoothed by kernel regression in [Cu05] Figure 6, solid line. A similar function appears in [Ko05] Figure 2, solid line. We stress that the same luck distribution is used for all funds.

The same bootstrap procedure also leads to a luck distribution PDF for the return of each fund after ranking by size of return. Examples of such distributions are given in [Cu05] Figure 5 (Note that, if there were 1000 funds in the sample the 99th percentile fund would be have rank 10 (or 11) from the top.) Similar displays are found in [Ko05] Figure 1. To avoid confusion we shall call this second type of distribution the 'luck distribution for order (rank) k'.

The conclusions of both papers follow from an analysis of the actual fund returns and the above luck distributions. There are two types of inference.

#### SKILL IN A SET OF FUNDS

As an example of this approach we point to Quote 6 of KTWW. Above a specified return R (10% in the quote), the area under the luck PDF corresponds to say NL (9) funds when the total number of funds in the sample is taken into account. In the distribution of actual fund returns, above R there are say NA (29) funds. If NA is substantially greater than NL then it is likely that some of the actual funds employed skill, in fact about NA - NL of them. However, we do not know which of the actual funds in that group had the skill.

We have no objection to the concept of this approach.

#### SKILL IN INDIVIDUAL FUNDS

Quote 1 of CNO, and other quotes, make it clear that CNO believe that they can improve on the above approach by identifying individual funds that have or do not have skill. Suppose that the order of the actual return of a fund is k, i.e. the return is k from the top. Let pv be the area of the upper tail of the luck distribution of order k above the actual return of the fund of order k. Choose a significance level, say 0.05. If pv < 0.05 then CNO claim that we can reject the null hypothesis that the performance of the fund of order k is attributable to luck. Instead the fund has genuine skill, they say.

For example [Cu05] Table 2, Panel A, Row p-tstat gives a value pv = 0.038 for the fund of order 10. Thus CNO declare that fund must have exhibited skill. However the fund of order 5, with pv = 0.157 showed no evidence of skill.

Since the order (rank) of the actual return depends on more than just the return of the fund of rank k, we question the validity of the CNO conclusions. This point is discussed below in the section entitled UNSOUND BASIS OF THE CNO TEST FOR SKILL. Tests on our model confirm this skepticism.

#### MODEL

In the model the observed fund return  $r_k$  is assumed to be drawn from a random variable  $\tilde{r}_k$ . We assume that  $\tilde{r}_k$  may be written as

$$\widetilde{r}_k = \widetilde{\rho} + m_k, \, k = 1, \dots, n \,, \tag{1}$$

where  $\tilde{\rho}$  is a mean-zero random variable, independent of k, and  $m_k$  is the mean of  $\tilde{r}_k$ . We interpret  $m_k$  to be a measure of the skill applied by the fund manager. A value  $m_k = 0$  means no skill, and the sign of  $m_k$  may be positive or negative, with high value meaning high skill.

It is clear that the PDF f(x) for  $\tilde{\rho}$  corresponds to the luck distribution as defined by KTWW and CNO. For our model we choose a particular form for f(x), with a corresponding cumulative distribution function (CDF) F(x) given by

$$F(x) = \int_{-\infty}^{x} dt f(t) \,. \tag{2}$$

Any form would do, but we use the standard normal distribution for ease of computation. We doubt whether this choice affects the general thrust of our results.

In common with the articles we assume that the observed set of fund returns  $r_k, k = 1, ..., n$  is obtained by drawing one from each of the random variables  $\tilde{r}_k$ . The articles contemplate the possibility that these draws are not independent, but for simplicity we assume independence of the *n* draws. The CNO technique should apply to this case.

The method of CNO is based on the notion of order statistics [Da03] for draws from the luck distribution, which has density f(x). Suppose that we draw *n* times to obtain  $X_1, X_2, \ldots, X_n$ , and that, ordered in decreasing size, these quantities are  $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ .

There is a CDF  $F_{(k)}(x)$  for each quantity  $X_{(k)}$ . For our model this is the CDF of the luck distribution for order k as used by KTWW and CNO. With the assumption of independence of the *n* draws, it is possible to relate  $F_{(k)}(x)$  to F(x) as follows [Da03]. In our notation

$$F_{(k)}(x) = \left[F(x)\right]^{n-k+1} \sum_{j=0}^{k-1} {\binom{n-k+j}{j}} (1-F(x))^j .$$
(3)

It follows that the probability that  $X_{(k)}$  is greater than x is given by

$$P(X_{(k)} > x) = 1 - F_{(k)}(x) = 1 - z^{n-k+1} \sum_{j=0}^{k-1} {\binom{n-k+j}{j}} (1-z)^j$$
(4)

with z = F(x).

The articles refer to  $P(X_{(k)} > x)$  as the p-value or p-stat (for order k and return x) - we called it pv above.

#### **CNO TEST FOR SKILL IN INDIVIDUAL FUNDS**

In testing for funds that perform better than by chance CNO choose a significance level with which to compare the p-value. Throughout the following we take that level to be 5%. For our model their procedure may be described by the following steps.

- 1. For any draw of *n* returns, including the observed data  $r_k$ , k = 1, ..., n, we order (rank) the returns, highest first, giving rise to a list  $r_{(k)}$ , k = 1, ..., n. Thus high order means low k.
- 2. For each k calculate the p-value  $p_k = P(X_{(k)} > r_{(k)})$  using (4).
- 3. Then the fund corresponding to order k in the list outperformed (involved skill, i.e. had a performance better than can reasonably be attributed to luck), if and only if  $p_k < 0.05$ .

#### **CNO TEST FOR SKILL IN THE MODEL**

It is relatively straightforward to apply the CNO test to the model described above. We concentrate on the upper tail of the distribution. Some helpful definitions are

• 
$$f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$$
 (5)

• 
$$F(x) = 0.5(1 + \operatorname{erf}(x/\sqrt{2}))$$
 (6)

• The binomial coefficient 
$$\binom{n-k+j}{j} = \frac{(n-k+j)!}{(j)!(n-k)!}$$
 (7)

We rearranged the CNO test by first finding, for a given n, the set of values  $z_k, k = 1, ..., K$  such that

$$=1-z_{k}^{n-k+1}\sum_{j=0}^{k-1} {\binom{n-k+j}{j}} (1-z_{k})^{j} = 0.05$$
(8)

In view of (4) the test then becomes

The fund at order k used skill if and only if 
$$F(r_{(k)}) > z_k, k = 1, ..., K$$
 (9)

For the test we chose n = 600 and K = 15. The values of  $z_k$  are given in the Table 1 below under the heading Order Points.

In the model we assume that a number of funds N have positive skill corresponding to a value of mean  $m_k = M$ . An equal number have negative skill with  $m_k = -M$ . The remaining funds have zero skill with mean zero. In the results listed in Table 2 we provide the data from simulations corresponding to N = 0, M = 0. and 16 other combinations of N and M, namely N = 10, (10), 40 and M = 0.5, (0.5), 2.0.

In each case we make 10,000 draws of 600 returns given by (1). We study only those returns that fall in the highest 15, corresponding to a total of 150,000 returns. In the Table 2 we provide for these returns the following totals for the 10,000 draws.

1. Real skill	All returns with positive skill by our definition
2. Apparent skill	All returns declared to have skill by the CNO method
3. Skill in apparent	All Type 2 returns with skill
4. No apparent skill	All returns declared to have no skill by the CNO method
5. Skill in no apparent skill	All Type 4 returns with skill

Note that totals obey the following relations .

Type 3 + Type 5 = Type 1 Type 2 + Type 4 = 150,000

We also break down Types 1, 2 and 3 returns according to the order involved, from 1 to 15. These are listed under the headings

- Real skill by order
- Apparent skill by order
- Real skill in apparent skill set.

In the rows designated A, B, C we provide measures of the success of the CNO method and two alternative methods of predicting which funds employed skill. The measures are

**A** CNO method It states that, if a fund return has order k, then skill was involved if (9) is satisfied. The measure is the percentage of funds with actual skill out of those with apparent skill at order k, i.e. the ratio of the totals for Type 3 to Type 2 expressed as a percentage.

**B** Unconditional high order method It assumes that high order is correlated with high skill. The measure is the ratio of the number of funds with real skill at order at least as high as k, out of all funds in the same set of orders.

**C** Conditional high order method If a the fund of order k has apparent skill (it satisfies (9)) then we predict that all funds of higher order will have skill. The measure is the ratio of the number of funds with real skill at order higher than k when the fund of order k satisfies (9), out of all funds in the same set of orders.

To illustrate, for example in the case with N = 20, M = 1.0 we find totals

30,497
20,552
4,961
129,448
25,536

For order 9 we find 1,453 funds declared to have skill by the CNO method, of which 286 actually have skill. For this order there are 1,793 funds in all that have skill, so that there are an additional 1,793 - 1,453 = 340 funds, not declared to have skill, that actually have skill.

Again for order 9, row A shows that the CNO method is correct  $20\% = 100 \times (286/1,453)$  of the time. Row B shows that choosing all funds with order 9 or higher would lead to a fund with skill 24% of the time. Row C shows that choosing all funds of order 8 or higher when the fund of order 9 satisfies the CNO condition (9) would lead to a fund with skill 27% of the time.

#### **RESULTS FROM THE MODEL**

The cases studied range from few managers with little skill to many managers with high skill. At the extreme of no managers with any skill, N = 0, M = 0.0, we obtain the expected result that each order should contain  $5\% \times 10,000 = 500$  funds that meet the condition (9), that is  $F(r_{(k)}) > z_k$ . Thus the numbers close to 500 listed under apparent skill by order represent the inevitable noise in measurement associated with the p-value of 5%. These results suggest that it is unlikely that there are large errors in the calculation of the order points and F(x).

At the other extreme of many managers with high skill, represented by N = 40, M = 2.0, it is not surprising that most of the top 15 funds have real skill, or that most of these funds satisfy the condition  $F(r_{(k)}) > z_k$ . Of almost 150,000 funds studied, nearly 120,000 have real skill, and almost all of these are declared by the CNO method to possess skill. However, the CNO method is inaccurate in wrongly declaring that another 30,000 funds will have skill. This number of false positives is much in excess of the inherent noise in the method mentioned above.

In between the two extremes the performance of the CNO method is worse. Often its choices contain a high percentage of false positives, but the method also fails to predict a significant number of funds that have real skill. For example, in the case N = 20, M = 1.5, the CNO method states that 56,653 funds will have skill, but only 22,260 of those in fact do so. Moreover, another 31,378 funds have skill not predicted by the CNO method.

The tables showing predicted and actual skill by order are interesting. There is a very clear trend for the percentage of funds of Type 3 (real skill in the CNO skill set) as a share of Type 2 (CNO predicted skill) to decrease with increasing order. For example, in the case N = 20, M = 1.0, row A shows that the percentage decreases steadily from 54% at order 1 to 15% at order 15. The methods of predicting skill summarized in rows B and C are superior to the CNO method of row A except for a few of the highest orders. The same conclusion applies to all the cases listed in Table 2.

CNO state that a fund of order k employed skill if and only if the condition  $F(r_{(k)}) > z_k$  holds. The above evidence supports the conclusion that the CNO statement is false. Of course, a single draw is all we have in the real world, so that the situation is even worse than that presented by the collected results from 10,000 draws.

#### **UNSOUND BASIS OF THE CNO TEST FOR SKILL**

The poor showing of the CNO test on our model suggests that we investigate the motivation behind the test. The key logic used by CNO reads as follows.

- Suppose that the actual return of the fund of rank k is  $r_{(k)}$ .
- Calculate the p-value  $p_k = P(X_{(k)} > r_{(k)})$  using (4).
- If  $p_k < 0.05$  we can reject the null hypothesis that the performance of the fund of order *k* is attributable to luck.

The validity of this reasoning is based on the belief that the p-value depends only on the return  $r_{(k)}$  and on no other actual fund return. This belief is incorrect, as we explain below.

In our model the process of drawing the 600 actual returns may be broken down into two stages.

- Make 600 independent draws from the distribution due to luck  $\tilde{\rho}$ , which has density f(x), and rank them highest return first.
- For those funds with skill, add the appropriate mean (M or -M) to the return drawn, and move it to its correct place in the ranking.

Consider the return at a given order k chosen in the range (1-15). It is likely (95%) that, before the addition of the means, the return at order k will not satisfy the CNO condition  $F(r_{(k)}) > z_k$ . By our assumption it is unlikely (20/600 in the above example) that the return at order k will possess skill, but, if it does, adding the mean leads to an increased probability that the CNO condition is satisfied. That situation would correspond to a case when the CNO method succeeds.

However, a more likely situation is when a fund with order lower than k (i.e. further from the top of the distribution) has skill, and the addition of M vaults the return to an order jabove k, i.e. j < k. This will mean that all funds (including fund k) previously at order j or lower will be pushed one step down the list. Thus the fund previously at order k will now have order k+1, which leads to an increased chance that it will now satisfy the CNO condition. Should that happen, we will have a fund now at order k+1 that will be predicted to have skill but in fact does not, i.e. a false positive from the CNO method.

Thus the null hypothesis is improperly formulated. It should relate to the performance of several funds, not just the one of order k.

# CONCLUSION

Our calculations on the model described above show conclusively that the CNO method does not identify funds with skill to anywhere near the accuracy expected. It makes errors of two kinds, picking some funds to have skill when they do not, and failing to pick some funds with skill. We have provided an explanation for why the method fails.

It should be noted that there are other aspects of the procedure of CNO, and indeed also of KTWW, that need careful scrutiny, namely

- The accuracy of the construction of probability distributions and order points near the extremes by bootstrap methods. This question could be investigated in our model;
- The use of factor models intended to describe certain types of risk. Daniel and Titman [Da05] have cast serious doubt on these procedures.

We stress that our arguments about the CNO method apply whether or not these additional concerns are justified.

## REFERENCES

Cu05 Cuthbertson, K., Nitzsche, D. and O'Sullivan, N., "Mutual Fund Performance: Skill or Luck?" Cass Business School Research Paper (2005) Available at SSRN: http://ssrn.com/abstract=665744

Da05 Daniel, K., and Titman, S., "Testing Factor-Model Explanations of Market Anomalies," Working paper (2005) http://www.kellogg.northwestern.edu/faculty/daniel/htm/

Da03 David, H., and Nagaraja, H., "Order Statistics," Third edition. Wiley (2003)

Ko05 Kosowski, R., Timmermann, A., Wermers, R. and White, H., "Can Mutual Fund 'Stars' Really Pick Stocks? New Evidence from a Bootstrap Analysis". (2005) Available at SSRN: http://ssrn.com/abstract=855425

# QUOTES

# **QUOTATIONS FROM KTWW [Ko05]**

## QUOTE 1 page 4 (in pdf) line 14

our bootstrap tests consistently indicate that the large positive alphas of the top ten percent of funds, net of costs, are extremely unlikely to be due to sampling variability (luck).

## QUOTE 2 page 5 line 6

The key to our study is the bootstrap analysis, which allows us to precisely separate luck from skill in the complicated non-normal cross-section of ranked mutual fund alphas.

#### QUOTE 3 page 17 line -7

our results of Panel A (Table II) show that funds with alphas ranked in the top decile generally exhibit significant bootstrapped p-values. However, this is not always the case. For example, the second-ranked fund ..... displays a large but insignificant alpha; this alpha simply is insufficiently large to reject (based on the empirical distribution of alphas) that the manager achieved it through luck alone.

#### QUOTE 4 page 19 line -1

the standard parametric p-value for the t-statistic (the one-tailed p-value for t=1.4 is roughly nine percent) indicates that the fund at the 10th - percentile exhibits a significant t-statistic, .... However the bootstrap does not find this t-statistic to be significant, and does not reject the null of no manager talent at the 10th - percentile (this p-value equals 25 percent).

## QUOTE 5 page 20 line 6

these observations reinforce our prior evidence that many superior and inferior funds exist in our sample. Since our interest is in the actual number of funds exceeding a certain level of alpha, compared to the bootstrapped distribution, we plot the CDF in Panel B.

## QUOTE 6 page 20 line 14

We can also use the bootstrapped distribution of alphas to calculate how many funds .... would be expected, by chance alone, to exceed a given level of performance ..... Panel A of Figure 3 indicates that nine funds should have an alpha estimate higher than 10 percent per year by chance - in reality 29 funds achieve this alpha.

# **QUOTATIONS FROM CNO [Cu05]**

# QUOTE 1 page 1 Abstract line 1

Using a comprehensive data set on ... UK equity mutual funds ... we use a bootstrap methodology to distinguish between 'skill' and 'luck' for *individual* funds.

# QUOTE 2 page 3 line 8

we use a cross-section bootstrap procedure across *all* individual funds. This enables our 'luck distribution' for any chosen fund (e.g. the best fund) to encapsulate possible outcomes of luck not just for our chosen fund but across all the funds in our data set. We are then able to separate 'skill' from 'luck' in performance tests of *individual* funds, even when the distribution of idiosyncratic risk across *many* funds is highly non-normal. This methodology has not been applied to UK data and was first applied to US mutual funds by Kosowski et al.

## QUOTE 3 page 5 line 23

.... of the top 20 ranked funds in the positive tail of the performance distribution, 7 funds exhibit levels of performance which cannot be attributable to 'luck' at 5% significance level

## QUOTE 4 page 12 line 9

Having obtained our 'luck distribution', we now compare the best fund's actual *ex-post* performance given by its estimated  $\hat{\alpha}_{max}$  against the 'luck distribution' for the best fund. If

 $\hat{\alpha}_{\max}$  exceeds the 5% right tail cut off point in  $f(\hat{\alpha}_{\max})$ , we can reject the null hypothesis that the performance of the best fund is attributable to luck. Above, we could have chosen any fund (e.g. the  $2^{nd}$  best fund) on which to base the 'luck distribution'. So, we can compare the actual *ex-post* ranking for any chosen fund against *its* luck distribution and separate luck from skill, for all *individual* funds in our sample.

# QUOTE 5 page 14 line 9

We can now compare any *ex-post*  $\hat{\alpha}_i$  with its appropriate 'luck distribution'. Suppose we are interested in whether the performance of the ex-post best fund is due to skill or luck. If  $\hat{\alpha}_{max}$  is greater than the 5% upper tail cut off point from  $f(\hat{\alpha}_{max})$  then we reject the null that its performance is due to luck (at 95% confidence). We infer that the fund has genuine skill. This can be repeated for any other point in the performance distribution, right down to the ex-post worst performing fund in the data.

# TABLE 1

Table 1. The order points  $z_k$  obtained by solving Eq. (8) for n = 600 funds and significance level 0.05.

ORDER	POINTS
1	0.99991451
2	0.99940741
3	0.99863584
4	0.99771969
5	0.99671084
6	0.99563628
7	0.99451203
8	0.99334851
9	0.99215294
10	0.99093055
11	0.98968526
12	0.98842015
13	0.98713762
14	0.98583964
15	0.98452783

# TABLE 2

Table 2. Results of model simulations for 17 cases specified by the number of funds with positive skill and the strength of that skill. The total number of funds was 600 and the significance level 0.05. The meaning of the data below is explained in the section entitled CNO TEST FOR SKILL IN THE MODEL.

14 15 0 0 507 510 0 0
14 15 0 0 507 510 0 0
0 0 507 510 0 0
507 510 0 0
0 0
0 0 0 0
0 0
14 15
422 422
600 591
28 19
5 3 5 5 5 5

А в С 

Real skill in apparent skill set 206 210 189 179 172 169 Α в 2.2 С