Fooled By Correlation: How Blind Acceptance of Correlation Dogma Destroys Diversification

Jeffry Haber
Iona College

Ask any quantitative finance academic whether you should use prices (NAVs) or returns when calculating correlations and they will tell you that you can ONLY use returns. Using an example of a hypothetical data set, the correlation of prices is 1.00 and the correlation of the returns derived from these prices is .02. The question for any thinking person is "looking at the two series of data, first for prices and then for returns, do the numbers look extremely correlated (prices) or do they look like they are nearly perfectly random between the two series (returns)?"

INTRODUCTION

The cornerstone of portfolio construction is diversification, and the quantification of diversification is correlation. Correlation has been a mainstay of the investing and financial toolbox ever since the early days of modern portfolio theory. Correlation can be computed using prices (net asset values or NAVs) of the investment, or the related return streams. Any quantitative financial professional will tell you that you can ONLY use returns because prices are not independent and not stationary.

If the correlations calculated using prices and returns were similar, this would not be an issue. If the correlations are widely different, then it could be a serious issue. Some investments are added to portfolios based on their ability to act as a diversifier, and the statistic used to make this determination is correlation. If the correlation calculation did not accurately reflect the effect the investment would have on the portfolio then the investor could be making a mistake.

Given that diversification is important and that correlation is the measurement we have to work with, the question becomes "can there be a difference between the correlation that is calculated using the prices and the correlation calculated using returns, and if so, which is the superior basis to use?"

RESEARCH METHODOLOGY

This paper develops two streams of data for 120 observations, representing two investment options and the ending monthly prices for each. This is shown in Figure 1. Of the 120 observations, 119 are exactly the same between the two investment options (labeled A and B). The lone difference occurs in observation 90. Since 119 observations out of 120 are exactly the same, it is not surprising that the correlation between the two streams is 1.00.

Next, I derive the returns from these streams of prices. This is shown in Figure 2. This yields 119 observations. Of the 119 observations, 117 are the same. The differences occur in observations 89 and 90. What is the resulting correlation of these 119 observations, where 117 are exactly the same? Nearly

perfect NON-correlation (almost 0.00). This is the polar opposite from the correlation that was calculated using prices.

WHICH IS CORRECT?

We have answered the first question, can the correlations calculated using prices and returns from the same data set produce different results. The answer is yes, and as different as possible. The next question is "which is the superior correlation to use."

I submit that a reasonable person looking at both sets of data would come away feeling that the streams of data between both investment options (A and B) are very similar in the case of prices AND in the instance of returns. Further, I propose that a reasonable person would say that both situations represent highly correlated data sets.

On this basis, given this set of data, the prices are clearly the superior basis for calculating correlation.

LIMITATIONS

The data set used was not actual data. The problem in using actual data is in answering the question of which correlation calculation is superior is that real data is very messy – a reasonable person usually cannot look at the streams and come to a reasoned conclusion on which is superior. The advantage of this contrived data set is that a reasonable person can look at the data set and come to a conclusion on what they expect the correlation calculation to be.

CONCLUSIONS

The first conclusion I would draw is that it is worthwhile to compute correlation using both prices and returns. Where they are similar there is no difficult decision to be made. Where they are starkly different some thought has to go into in deciding which to utilize. Each investor will have to make their own decision about the tie-breaking procedure they will use.

The second conclusion is that the statistical benefits that come from using returns instead of prices (such as independence and stationarity) ignore that returns are derived from prices. There is no magic (aside from statistical) that comes from translating a price into a return using the formula:

FORMULA 1
FORMULA FOR RETURN

$$R_2 = P_2 \_ P_1 / P_1$$

Where R = Return, P = Price

The third conclusion is that this data set used streams of prices that were very consistent (little volatility). The findings may not apply to data that is more volatile.

Lastly, and perhaps most importantly, sometimes you have buck the majority and think for yourself. Just because some/many/most feel that there is only one way to do something, such as using returns in correlation calculations, always come to an independent decision. Don't follow the consensus when data says otherwise.

FIGURE 1
CORRELATION OF PRICES (NAVs)

Obs	A	B
1	100	100
2	105	105
3	110	110
4	116	116
5	122	122
6	128	128
7	134	134
8	141	141
9	148	148
10	155	155
11	163	163
12	171	171
13	180	180
14	189	189
15	198	198
16	208	208
17	218	218
18	229	229
19	241	241
20	253	253
21	265	265
22	279	279
23	293	293
24	307	307
25	323	323
26	339	339
27	356	356
28	373	373
29	392	392
30	412	412
31	432	432
32	454	454
33	476	476
34	500	500
35	525	525
36	552	552
37	579	579
38	608	608
39	639	639
40	670	670

Obs	A	B
41	704	704
42	739	739
43	776	776
44	815	815
45	856	856
46	899	899
47	943	943
48	991	991
49	1,040	1,040
50	1,092	1,092
51	1,147	1,147
52	1,204	1,204
53	1,264	1,264
54	1,327	1,327
55	1,394	1,394
56	1,464	1,464
57	1,537	1,537
58	1,614	1,614
59	1,694	1,694
60	1,779	1,779
61	1,868	1,868
62	1,961	1,961
63	2,059	2,059
64	2,162	2,162
65	2,270	2,270
66	2,384	2,384
67	2,503	2,503
68	2,628	2,628
69	2,760	2,760
70	2,898	2,898
71	3,043	3,043
72	3,195	3,195
73	3,355	3,355
74	3,522	3,522
75	3,698	3,698
76	3,883	3,883
77	4,077	4,077
78	4,281	4,281
79	4,495	4,495
80	4,720	4,720

Obs	A	B
81	4,956	4,956
82	5,204	5,204
83	5,464	5,464
84	5,737	5,737
85	6,024	6,024
86	6,325	6,325
87	6,642	6,642
88	6,974	6,974
89	7,322	7,322
90	7,689	4,900
91	8,073	8,073
92	8,477	8,477
93	8,900	8,900
94	9,346	9,346
95	9,813	9,813
96	10,303	10,303
97	10,818	10,819
98	11,360	11,360
99	11,927	11,927
100	12,524	12,524
101	13,150	13,150
102	13,808	13,808
103	14,498	14,498
104	15,223	15,223
105	15,984	15,984
106	16,783	16,783
107	17,622	17,622
108	18,503	18,503
109	19,429	19,429
110	20,400	20,400
111	21,420	21,420
112	22,491	22,491
113	23,616	23,616
114	24,796	24,796
115	26,036	26,036
116	27,338	27,338
117	28,705	28,705
118	30,140	30,140
119	31,647	31,647
120	33,230	33,230

FIGURE 2
CORRELATION OF RETURNS

Obs	A	B
1	0.05	0.05
2	0.05	0.05
3	0.05	0.05
4	0.05	0.05
5	0.05	0.05
6	0.05	0.05
7	0.05	0.05
8	0.05	0.05
9	0.05	0.05
10	0.05	0.05
11	0.05	0.05
12	0.05	0.05
13	0.05	0.05
14	0.05	0.05
15	0.05	0.05
16	0.05	0.05
17	0.05	0.05
18	0.05	0.05
19	0.05	0.05
20	0.05	0.05
21	0.05	0.05
22	0.05	0.05
23	0.05	0.05
24	0.05	0.05
25	0.05	0.05
26	0.05	0.05
27	0.05	0.05
28	0.05	0.05
29	0.05	0.05
30	0.05	0.05
31	0.05	0.05
32	0.05	0.05
33	0.05	0.05
34	0.05	0.05
35	0.05	0.05
36	0.05	0.05
37	0.05	0.05
38	0.05	0.05
39	0.05	0.05
40	0.05	0.05
41	0.05	0.05
42	0.05	0.05
43	0.05	0.05
44	0.05	0.05
45	0.05	0.05
46	0.05	0.05
47	0.05	0.05
48	0.05	0.05
49	0.05	0.05
50	0.05	0.05
51	0.05	0.05
52	0.05	0.05
53	0.05	0.05
54	0.05	0.05
55	0.05	0.05
56	0.05	0.05
57	0.05	0.05
58	0.05	0.05
59	0.05	0.05
60	0.05	0.05
61	0.05	0.05
62	0.05	0.05
63	0.05	0.05
64	0.05	0.05
65	0.05	0.05
66	0.05	0.05
67	0.05	0.05
68	0.05	0.05
69	0.05	0.05
70	0.05	0.05
71	0.05	0.05
72	0.05	0.05
73	0.05	0.05
74	0.05	0.05
75	0.05	0.05
76	0.05	0.05
77	0.05	0.05
78	0.05	0.05
79	0.05	0.05
80	0.05	0.05
81	0.05	0.05
82	0.05	0.05
83	0.05	0.05
84	0.05	0.05
85	0.05	0.05
86	0.05	0.05
87	0.05	0.05
88	0.05	0.05
89	0.05	-0.33
90	0.05	0.65
91	0.05	0.05
92	0.05	0.05
93	0.05	0.05
94	0.05	0.05
95	0.05	0.05
96	0.05	0.05
97	0.05	0.05
98	0.05	0.05
99	0.05	0.05
100	0.05	0.05
101	0.05	0.05
102	0.05	0.05
103	0.05	0.05
104	0.05	0.05
105	0.05	0.05
106	0.05	0.05
107	0.05	0.05
108	0.05	0.05
109	0.05	0.05
110	0.05	0.05
111	0.05	0.05
112	0.05	0.05
113	0.05	0.05
114	0.05	0.05
115	0.05	0.05
116	0.05	0.05
117	0.05	0.05
118	0.05	0.05
119	0.05	0.05