NCCC-134

APPLIED COMMODITY PRICE ANALYSIS, FORECASTING AND MARKET RISK MANAGEMENT

Performance Persistence for Managed Futures

by

B. Wade Broersen and John Townsend

Performance Persistence for Managed Futures

B. Wade Brorsen and John Townsend[*]

Past literature on managed futures funds has found little evidence that the top performing funds can be predicted. But, the past literature has used small datasets and methods which had little power to reject the null hypothesis of no performance persistence. The objective of this research is to: determine whether performance persists for managed futures advisors using large datasets and methods which have power to reject the null hypothesis.

We use data from public funds, private funds, and commodity trading advisors (CTAs). The analysis proceeds in four steps. First, a regression approach is used to determine whether after adjusting for changes in overall returns and differences in leverage that funds all have the same mean returns. Second, we use Monte Carlo methods to demonstrate that Elton, Gruber, and Rentzler's methods have little power to reject false null hypotheses and will reject true null hypotheses too often. Third, we conduct an out-of-sample test of various methods of selecting the top funds. Fourth, since we do find some performance persistence, we seek to explain the sources of this performance persistence by using regressions of (a) returns against CTA characteristics, (b) return risk against CTA characteristics, (c) returns against lagged returns, and (d) changes in investment against lagged returns.

Performance persistence could exist due to either differences in cost or differences in the skill of the manager. Our results favor skill as the explanation since returns were positively correlated with cost. The performance persistence is statistically significant, but is small relative to the variation in the data (only 2-4% of the total variation). But, the performance persistence is large relative to the mean. Monte Carlo methods showed that the methods used in past research could often not reject false null hypotheses and would reject true null hypotheses too often.

Introduction

Schwager in his book *Managed Trading: Myths & Truths* reviews the literature on whether managed futures funds exhibit performance persistence and conducts his own analysis. He concludes that there is little evidence that the top performing funds can be predicted. But, the entire set of literature has used variations of the methods of Elton, Gruber, and Rentzler (EGR). Grossman in an unpublished 1987 article demonstrated that EGR's methods had virtually no power to reject the null hypothesis of no predictability (when the assumptions of EGR's methods are true). EGR's and Schwager's methods do not fully consider the problems created by differences in leverage. EGR argued incorrectly that their methods are appropriate because of

cross-sectional nonnormality, but the real statistical problem is heteroskedasticity. Regression is a far more powerful procedure than the nonparametric methods used in past literature. The objective of the research project discussed in this report is therefore: determine whether performance persists for managed futures advisors using methods which properly account for variation in overall fund returns and for heteroskedasticity.[1]

We use data from public funds, private funds, and commodity trading advisors (CTAs). The analysis proceeds in four steps. First, a regression approach is used to determine whether after adjusting for changes in overall returns and differences in leverage that funds all have the same mean returns. Second, we use Monte Carlo methods to demonstrate that EGR's methods have little power to reject false null hypotheses and will reject true null hypotheses too often. Third, we conduct out-of-sample tests similar to those of EGR and Schwager. Fourth, since we do find some performance persistence, we seek to explain the sources of this performance persistence by using regressions of (a) returns against CTA characteristics, (b) return risk against CTA characteristics, (c) returns against lagged returns, and (d) changes in investment against lagged returns.

Data

The data were provided by LaPorte Asset Allocation. Much of the data originated from *Managed Accounts Reports*, but some were from other sources. The CTA data include information on CTAs no longer trading as well as CTAs who are still trading. The data include monthly returns beginning in 1978 and ending in May 1996. The data set as provided recorded zeroes when an observation was missing. Missing values were deleted by deleting observations where returns and net asset value were zero. This should help prevent deleting observations where returns were truly zero. The return data were converted to log changes[2], so they should be interpreted as percentage changes in continuous time.

The mean returns are presented in Table 1. As in past literature, the CTA data have the highest mean returns. The CTA returns may be higher because of selectivity biases. Selectivity bias is not a major concern here, since we are comparing CTAs to other CTAs and not to some other investment. The data show considerable kurtosis (table 1). However, as we will demonstrate, this kurtosis is caused by heteroskedasticity (returns of some funds are more variable than others).

---

[1]Heteroskedasticity means that some observations have higher variances than others.

[2]The formula used was $r_{it} = (1 + d_{it}/100) * 100$ where d_{it} is the discrete time return. The adjustment factor of 100 is used since the data are measured as percentages.

Table 1. Descriptive Statistics for the Public, Private, and Combined CTA Data Sets and Continuous Time Returns

Data set	Public funds	Private funds	Combined CTA's
Observations	32420	23723	57018
# Funds	577	435	1071
Percentage Returns			
Mean	0.31	0.62	1.28
SD	7.68	9.22	10.53
Minimum	-232.69	-224.81	-135.48
Maximum	229.73	188.93	239.79
Skewness	-2.08	-0.49	1.14
Kurtosis	133.91	40.70	24.34

Regression Test of Performance Persistence

To measure performance persistence, a model of the stochastic process that generates returns is required. The process considered is

$$(1) \qquad r_{it} = \alpha_i + \beta_i \bar{r}_t + \epsilon_{it} \quad i = 1, ..., n \qquad t = 1, ..., T$$

$$\epsilon_{it} \sim N(0, \sigma_i^2)$$

where r_{it} is return of fund (or CTA) i in month t, $\bar{r}_t$ is average fund returns in month t, and the parameter β_i reflects differences in leverage. The model allows each fund to have a different

Only funds/CTAs with at least three observations are included. The model is estimated using feasible generalized least squares by using PROC GLM of SAS.

The null hypothesis, to be considered is whether once adjusted for changes in overall returns and differences in leverage, do funds all have the same mean returns? This is equivalent to testing the null hypothesis H_0: $\alpha_i = \bar{\alpha}, \forall i$ where $\bar{\alpha}$ is an unknown constant.

The results in Table 2, consistently show that funds and pools do not all have the same mean returns. This finding does contrast with previous research, but it is not really surprising given that funds and pools have different costs and that their trading systems and commodities traded vary widely.

Only about 2-4% of the variation in monthly returns across funds can be explained by differences in individual means. Since the predictable portion is so small, precise methods are needed to find it. Without the correction for heteroskedasticity and with public pool data the null hypothesis would not have been rejected. Even though the predictability is low, it is economically significant. The standard deviations in Table 2 are large and so 2-4% of the standard deviation is about 50% of the mean. Thus, even though there is a lot of noise, there is still potential to use past returns to predict future returns.

As shown in Table 3, the null hypothesis that each fund has the same variance was rejected. This is consistent with previous research which shows that some funds or CTAs consistently have more variable returns than others. The skewness in the residuals is removed and the kurtosis is greatly reduced. This demonstrates that most but not all of the nonnormality shown in Table 1 is really due to heteroskedasticity.

Table 2. Weighted ANOVA Table: Returns Regression for Public Funds, Private Funds, and Combined CTA Data

Fund Type	Public funds	Private funds	Combined CTA's
Sum of Squared Errors			
Ind. means	1751	1948	2333
Group mean	28335	10882	22751
Corrected Total	62221	36375	82408
R^2	0.48	0.35	0.31
Mean			
α	0.278	0.297	1.099
Variance			
α	1.16	2.277	2.240
F-statistic			
α's	2.94	4.32	2.12
β's	47.44	24.10	20.61

Table 3. F-Statistics for the Test of Homoskedasticity Assumption and Moments of Normality of Rescaled Residuals

Data set	Public funds	Private funds	Combined CTA's
Homoskedasticity	1.41		5.15
Residuals			
Skewness	-0.17	-0.02	0.35
Relative Kurtosis	3.84	3.05	2.72

Monte Carlo Study

The method used by EGR and other past research was to rank the funds by mean return or modified Sharpe ratio (mean divided by standard deviation) in one period. They then looked at whether the funds that ranked high in the first period also ranked high in a second period. We use Monte Carlo simulation to determine the power and size of hypothesis tests with EGR's methods when data follow the process assumed in (1). Data sets were generated by specifying values of α, β, and σ. The simulation used one thousand replications and 120 simulated funds. The mean return over all funds $\bar{r}_t$, is derived from the values of α and β as

$$(2) \qquad \bar{r}_t = \frac{\dfrac{\sum \alpha_i}{nfunds} + \dfrac{\sum \epsilon_{it}}{nfunds}}{1 - \dfrac{\sum \beta_i}{nfunds}}$$

The data sets were generated using the Interactive Matrix Language (IML) module of SAS. Data sets were generated using a fixed value of α to simulate no performance persistence. For the data sets generated with persistence present, α was varied and using the approximate distribution of α's in each of the three datasets. For the fixed value, the β's were set to a value of .5 For the generated data sets with differing leverage amounts, the β's were .5, 1, 1.5, and 2.

EGR assumed homoskedasticity and therefore data sets were generated with a fixed value of 2. Heteroskedasticity was created by letting the values of σ be 5, 10, 15, and 20 with one-fourth of the observations using each value. This allowed comparing the Spearman coefficient calculated for data sets with homoskedasticity and data sets with heteroskedasticity.

The funds were ranked in ascending order of returns for period one (first 12 months) and period 2 (last 12 months). From each 24 month period of generated returns, Spearman coefficients were calculated to identify the correlation between a fund's rank in period one and period two. For the Spearman coefficient, if the number of pairs of rankings is greater than 10, the distribution can be approximated by a normal. Since, 120 pairs are used in the simulation, the normal distribution is used.

Mean returns were also calculated for each fund in period one and period two and ranked. The funds were divided into groups consisting of the top third highest mean returns, middle third mean returns, and bottom third mean returns. Two additional subgroups were separated out, the top 3 highest mean returns funds and the bottom three funds with the lowest mean returns. The means across all funds in the top third group and bottom third group were calculated.

The results in Table 4 show what happens when EGR's test is used and there is no performance persistence. If the test is working correctly, the fail to reject probability should be

.95. When heteroskedasticity is present (data generation methods 2 and 3), too many rejections occur and the probability of not rejecting is less than .95. The heteroskedasticity may be more extreme in actual data, so the problem with real data may be worse than the rather small number of excess rejections found here.

Table 4. EGR Performance Persistence Results from Monte Carlo Generated Data Sets: No Persistence Present by Fixing α's to 1

Generated Data Subgroups	Data Generation method		
	1[a]	2[b]	3[c]
Mean returns			
top 1/3	1.25	1.25	0.70
middle 1/3	1.25	1.25	0.72
bottom 1/3	1.25	1.22	0.68
top 3	1.25	1.15	0.61
bottom 3	1.26	1.19	0.68
p-values			
reject-positive z	.021	.041	.041
reject-negative z	.028	.037	.039
fail to reject	.951	.922	.920
test of 2 means			
reject-positive	.026	.032	.032
reject-negative	.028	.020	.026
fail to reject	.946	.948	.942

[a]Data generated using $\alpha=1$, $\beta=.5$, $\sigma=2$.
[b]Data generated using $\alpha=1$, $\beta=.5$, $\sigma=5, 10, 15, 20$.
[c]Data generated using $\alpha=1$, $\beta=.5, 1, 1.5, 1$, $\sigma=5, 10, 15, 20$.

Next, we look at the probability of EGR's method finding performance persistence when performance persistence really exists (table 5). In this case, the fail to reject probability should be close to zero. The Spearman correlation coefficients show some ability to detect persistence with the large differences found in CTA data. But, they show little ability to find persistence at the levels in the public fund data used by EGR. The test of two means has even less ability to detect persistence. Thus, while the results do indicate weaknesses of EGRs methods, they do not appear as weak as we had expected.

Table 5. EGR Performance Persistence Results from Monte Carlo Generated Data Sets: Persistence Present by Allowing α's to Vary

Generated Data Subgroups	Data Generation method			
	1[a]	2[b]	3[c]	4[d]
Mean returns				
top 1/3	3.21	2.77	2.57	1.48
middle 1/3	1.87	2.09	1.85	1.30
bottom 1/3	0.80	1.41	1.15	1.14
top 3	4.93	3.47	3.26	1.68
bottom 3	-1.60	1.14	0.86	1.06
p-values				
reject-positive z	1.000	.827	.823	.149
reject-negative z	0.00	.000	.000	.003
fail to reject	.000	.173	.177	.848
test of 2 means				
reject-positive	1.00	.268	.258	.043
reject-negative	.000	.000	.000	.012
fail to reject	.000	.732	.742	.945

[a]Data generated using $\alpha=N(1.099,4.99)$, $\beta=.5, 1, 1.5, 2$, $\sigma=2$.
[b]Data generated using $\alpha=N(1.099,4.99)$, $\beta=.5$, $\sigma=5, 10, 15, 20$.
[c] Data generated using $\alpha=N(1.099,4.99)$, $\beta=.5, 1, 1.5, 2$, $\sigma=5, 10, 15, 20$
[d] Data generated using $\alpha=N(1.099,1)$, $\beta=.5, 1, 1.5, 2$, $\sigma=5, 10, 15, 20$.

Historical Performance as an Indicator of Later Returns

We now provide results based on methods similar to those of EGR. The previous Monte Carlo findings were based on a 1-year selection period and a 1-year performance period. Given the low power of EGR's method we use longer periods here. We use (i) a 4-year selection period with a 1-year performance period and (ii) a 3-year selection period with a 3-year performance period. Equation (1) was estimated for the selection period and the performance period. Since the returns are monthly, funds that had fewer than 60 or 72 monthly observations respectively were deleted to avoid having missing months of data.

The first 5-year period evaluated was 1980-84. The next five-year period was 1981-85. We consider three methods of ranking the funds: the α 's (intercept), the mean returns, and α/σ. For each parameter derived from the regression, a Spearman coefficient was calculated between the rank of the performance measure in the selection period and the rank with the measure for the performance period. The same statistic was used to rank funds in both periods. The null hypothesis of no correlation between ranks is tested with a z test. Because of the smaller sample

used here and the less efficient nonparametric method, this approach is expected to have less power than the direct regression test in (1).

A summary of the out-of-sample testing is reported in Table 6. The results for each year are reported in the appendix (Tables 12-17). The results in Table 6 are a summary of the annual results. Because of the overlap, the correlations from different time periods are not independent, so some care is needed in interpreting the results.

Regardless of the measure used, there is some positive correlation indicating performance persistence. The correlations are small which is consistent with the regression results. While there is performance persistence, it can be hard to find because of all the other random factors influencing returns.

The return/risk measure (α/σ) clearly shows the most performance persistence. Rankings based on mean returns are similar to the rankings based on α's. As the appendix shows, their correlations are similar in each year. Therefore, there does not appear to be as much gain as expected in adjusting for the overall level of returns.

The three-year selection period and three-year trading period show even higher correlations except for the early years of public funds. There were few funds in these early years and so their correlations may not be estimated very accurately. The rankings in the three-year performance period would have less variability than the one-year performance period. The higher correlation with the longer trading period suggests that performance persistence continues for a long time. This suggests that investors would want to be slow to change their allocations among managers.

The next question is why do these results seem to differ from past research? EGR actually did find similar levels of performance persistence, they just dismissed it as being small and statistically insignificant. The larger sample available now leads to more powerful tests. McCarthy did find performance persistence, but his results were discounted because his sample size was small. Irwin, Krukmeyer, and Zulauf used an alternative method of putting funds into quintiles. Their approach is difficult to interpret and may have led to low power. It is hard to say whether they found negative or positive correlation. Schwager did find a similar correlation of .07 for mean returns. Schwager, however, found a negative correlation for his return/risk measure. Schwager ranked funds based on return/risk when returns were positive, but ranked on returns only when returns were negative. This hybrid measure may have led to the negative correlation. Therefore, past literature really is consistent with a small amount of performance persistence. We were able to find the performance persistence because of the larger sample size and some slight improvements in methods. As shown in Table 6, several years did yield negative correlations and many positive correlations were statistically insignificant. Therefore, results over short time periods will be erratic.

Table 6. Summary of Spearman Correlations between Selection and Performance Periods.

Dataset selection criterion	Average correlation	Years positive (%)	Years positive and significant (%)
Four and one[a]			
CTA			
mean returns	.118	83	25
α	.114	83	25
α/σ	.168	100	42
Public funds			
mean returns	.084	75	33
α	.088	75	33
α/σ	.202	83	42
Private funds			
mean returns	.068	58	17
α	.047	58	0
α/σ	.322	92	50
Three and Three[b]			
CTA			
mean returns	.188	91	55
α	.186	91	45
α/σ	.253	100	64
Public funds			
Mean returns	-.015	45	36
α	.001	45	36
α/σ	.149	55	36
Private funds			
Mean returns	.212	91	36
α	.221	91	36
α/σ	.405	100	64

[a]Correlation between a four-year selection period and a one-year performance period. Averages are across the twelve one-year performance periods The same statistic was used for the rankings in each period.
[b]Three-year selection period and three-year trading period.

There is no strong difference in performance persistence between CTAs, public funds, and private funds. The performance persistence could be due to either differences in trading skill or differences in costs.

Performance Persistence and CTA Characteristics

Since some performance persistence was found, we next look at whether the differences can be explained. The monthly percentage returns were regressed against a set of CTA characteristics.

The means of the CTA characteristics are in table 7. In addition to the variables listed in the table, dummy variables were included for whether a long-term or medium-term trading system was used. Only the variables for dollars under management and time in existence change over time.

The data as provided by LaPorte Asset Allocation required some assumptions be made. If commissions, administrative fees, and incentive fees were all listed as zero, the observations for that CTA were deleted. This eliminated most, but not all of the missing values. If commissions were zero, the mean of the remaining observations was used.

A few times, options or interbank percentages were entered only as a yes. In these cases, the mean of the other observations using options or interbank were used. When no value was included for Non-U.S., Options, or Interbank, these variables were given a value of zero. Margins were often entered as a range. In this case, the midpoint of the range was used. When only a maximum was listed, the maximum was used.

For the trading horizon, if both short and medium term were listed, the horizon was classed as short term. If both medium and long-term or all three were listed, it was classed as medium term. Any observations with dollars under management equal zero were deleted.

The database contained a few additional variables. These included number of round turns, number of staff, and several verbal descriptions of the trading systems used. The number of round turns per dollar traded was considered as an explanatory variable, but was dropped since it measures the same thing as Margin and it had more missing observations. The number of staff had an insignificant negative coefficient and was dropped from the final model since there was no clear theoretical reason for including it. Attempts were made to form variables from the verbal descriptions of the trading system such as whether the phrase trend following was included. No significance was found. These variables are not included in the reported model since many descriptions were incomplete. Thus, the insignificance of the trading system could be due to the errors in the data. The remaining data are likely not error free. The most likely source of error would be treating a missing value as a zero. The data are originally from a survey so the survey itself could have had some errors. The presence of random errors in the data would cause the coefficients to be biased toward zero. Thus, one needs to be especially careful in not interpreting an insignificant coefficient as being zero. The data are for recent time periods. For example, the fees charged are the most recent fees. The fees are about half of what Irwin and Brorsen reported for public funds in the early 1980s.

Table 7. Mean and Standard Deviation of CTA Characteristics

Variable	Units	Mean	SD
Commission	percent of equity	5.7	4.7
Administrative fee	percent of equity	2.5	1.5
Incentive fee	percent of profits	19.9	4.5
Discretion	percent	27.7	37.9
Non-U.S.	percent	17.0	26.3
Options	percent	5.3	15.7
Interbank	percent	13.9	29.3
Margin	percent of equity invested	21.8	10.9
Time in existence	months	55.0	45.4
First year		87.9	4.9
Dollars under management ($million)		34.8	131.6

Note: These statistics are calculated using the monthly data so these statistics are weighted by the number of returns in the dataset.

The estimation method corrected for heteroskedasticity in that every CTA was allowed to have a different variance. Random effects were included for time and for CTA. The conclusions were unchanged when fixed effects were used for time. Considering random effects for CTAs is important since many of the variables do not vary over time. Ignoring random effects could cause significance levels to be overstated. Since the model is unbalanced (some CTAs have more monthly returns than others), PROC MIXED of SAS was used to estimate the parameters.

The results of regressing monthly percentage returns against CTA characteristics are presented in table 8. The short-term horizon traders have been outperformed by the long-term and medium-term traders. The coefficient of 0.30 for medium-term traders means that monthly percentage returns are 0.30 higher for medium-term traders than for short-term traders. For comparison, CTA monthly returns averaged 1.28 percent.

All three fee variables had positive coefficients with two of them statistically significant. The fee variables represent the most recent fees. This means that CTAs with larger historical returns charge higher fees. It may also mean that CTAs with superior ability are able to charge a higher price. A twenty percent incentive fee corresponds to monthly returns 0.44 percentage points higher than a CTA with no incentive fee so the coefficient estimates could be considered large.

None of the coefficients for Discretion, Non-U.S., Options, Interbank, and Margin were statistically significant. The set of dummy variables for commodities traded was also not statistically significant. The coefficients for Options and Interbank though cannot be considered

347

Table 8. Regressions of Monthly CTA Returns vs. Explanatory Variables

Variable	Coefficient	t-value
Intercept	13.90*	
Long-term	0.21*	
Medium-term	0.30**	
Commission	0.014	
Administrative fee	0.066**	
Incentive fee	0.022*	
Discretion	-0.001	
Non-U.S.	0.002	
Options	-0.004	
Interbank	0.003	
Margin	0.004	
Time in existence	-0.016**	
First year	-0.145*	
Dollars under management	-0.00104**	
F-test for commodity		0.51
F-test for time		9.05**
F-test of homoskedasticity		8.71**

small since these variables range from zero to one hundred. Thus, the coefficient of -0.004 means that firms with all trading in options are estimated to have returns .4 percentage points lower than a CTA that did not trade options.

Both the time in existence and the year trading began had negative coefficients. The negative sign is at least partly due to selectivity bias. Some CTAs were added to the database after they began trading. CTAs with poor performance may not have provided data. This could cause CTAs in their first years of trading to have higher returns. The first year coefficient's negative sign suggests that the firms entering the database in more recent years have lower returns. Thus, selectivity bias may be less in more recent years.

CTA returns may also genuinely erode over time. If CTAs do not change their trading system over time, others may discover the same inefficiency through their own testing. Also others may imitate the way the CTA trades by the CTA telling others about their system. CTAs are clearly concerned about this since most try to keep their system secret and have employees sign no compete agreements.

The dollars under management have a negative coefficient. The coefficient implies that for each million dollars under management, returns are 0.00104 percentage points lower. This

could occur due to increased liquidity costs from larger trade sizes. Returns would go to zero when a CTA had a billion dollars under management.

Following Goetzmann, Ingersoll, and Ross's arguments for hedge funds, managed futures exist because of inefficiencies in the market and because the CTA either faces capital constraints or is risk averse. By the very action of trading, the CTA is acting to remove these inefficiencies. Goetzmann, Ingersoll, and Ross argue that incentive fees exist partly to keep a manager from accepting too much investment. The dollars under management is a very crude measure of what is too much investment. Funds that trade more markets, more systems or less intensively could presumably handle more investment without decreasing returns.

We also estimated a similar model to explain the differences in the riskiness of the CTA returns (table 9). The most important factor determining the riskiness of CTAs is the percentage devoted to margins. While diversified funds were the least risky, the difference was not statistically significant. CTAs are becoming less risky. More recent CTAs have lower risk and CTAs have lowered their risk over time.

Commissions have a positive coefficient, but this may only mean that CTAs who trade larger positions generate more commissions. Incentive fees seem to encourage risk taking. Since the incentive fee is an implicit option, the CTA should earn higher incentive fees by adopting a more risky strategy. CTAs with more in Non-U.S. markets tend to have lower risk. Presumably the Non-U.S. markets provide some additional diversification.

Regressions of Returns Against Lagged Returns

To get some idea of the weights to put on various lags, monthly returns were regressed against average returns each of the last three years and the standard deviation of returns over the last three years. The model was estimated correcting for heteroskedasticity assuming each CTA had a different variance and with dummy variables for time. Ordinary least squares and random effects for time did not yield appreciably different results. Random or fixed effects for CTAs are not included since results using them with randomly generated data showed excessive statistical significance.

The results show some cycles in CTA and fund returns (table 10). CTAs tend to do well relative to other CTAs every other year. The sum of the three coefficients is positive which confirms the previous results regarding a small amount of performance persistence. The negative coefficient on lag 1-year returns supports Schwager's arguments that CTA/fund returns are negatively correlated in the short run.

Table 9. Regressions of Absolute Value of Residuals vs. CTA Characteristics

Variable	Coefficient	t-value
Long-term	0.027	0.06
Medium-term	0.083	0.24
Commission	0.117*	3.52
Administrative fee	-0.162	-1.37
Incentive fee	0.097*	2.29
Discretion	0.003	0.67
Non-U.S.	-0.013*	-2.39
Options	-0.011	-1.30
Interbank	-0.008	-1.02
Margin	0.092*	7.21
Time in existence	-0.029*	-10.45
First year	-0.260*	-5.34
Dollars under management	-0.001	-0.78
F-test for commodities traded	1.13	
F-test for time	7.74*	
F-test for homoskedasticity	11.96*	

Note: The absolute value of residuals is a measure of riskiness.

More risk as measured by historical standard deviation leads to higher returns for CTAs. Since CTAs are profitable, CTAs with higher leverage should make higher returns and have more risk. In contrast, both public and private funds have lower returns. Since their returns are low, leverage does not help as much in increasing returns. Also, multi-manager funds likely have lower risk. Multi-manager funds might have higher returns since they can adjust money between managers based on past performance.

The finding of negative correlations for the most recent years returns was investigated further. The most recent three months returns were separated. A slope dummy was added for whether the return was positive or negative. The results show little correlation for lagged positive returns, but losses tend to lead to higher than average gains two and three months later.

The results do offer some support for portfolio rebalancing and for Schwager's argument that investing with a manager after recent losses is a good idea. One difficulty of Schwager's arguments is why investing with poor performing managers would be a good idea. His argument is that returns are influenced by the amount of money devoted to a trading system. This idea is supported here by the results in table 11 and is also supported by Goetzmann, Ingersoll, and Ross. We now test whether money does flow out as Schwager suggested. We regressed the new

money in dollars under management (monthly percentage change in dollars - percent returns) vs. lagged returns and lagged standard deviations. The term new money may be a misnomer since money tends to be withdrawn rather than added. The lags for the most recent three months were separated and a dummy variable was added for positive returns.

Table 10. **Regressions of Monthly Managed Futures Returns against Lagged Returns and Lagged Sandard Deviation.**

Regressor	CTAs	Public	Private
Average returns 1-12 months ago	-.049* (-1.97)	-.059 (-2.45)	-.009 (-.33)
Average returns 13-24 months ago	.130* (5.93)	.160* (7.02)	.142* (5.46)
Average returns 25-26 months ago	.069* (3.53)	.074* (3.74)	.027 (1.33)
Standard-deviation last 3 years	.056* (4.16)	-.024 (-1.95)	-.027 (-1.86)
F-test of time fixed effects	35.38*	83.60*	28.29*

The results in Table 11 do show that investment and disinvestment are a function of lagged returns. Only returns in the most recent two years were statistically significant. Some asymmetry is found. The disinvestment due to negative returns is greater than the investment that occurs with positive returns for the most recent two months. There is no asymmetry for lag three months.

The flow of dollars does not match the changes in expected returns. People put most weight on the most recent past. There is some overreaction to short-run losses. The movement of money out of funds may explain at least part of the short-run negative autocorrelations in returns. Thus, the results do offer some support for Schwager's hypothesis.

Practical Implications

The research shows that some funds and CTAs have higher returns than others. Given the importance of the subject, we will address how to pick the best funds. First, remember that the performance persistence is small and that some years any method used will do worse than the average across all funds.

Table 11. Regression of Monthly CTA Returns and New Money Against Various Functions of Lagged Returns.

Variable	Monthly Returns	New Money[a]
One month ago returns	0.001	0.155*
	(0.04)	(5.94)
One month ago gains	0.026	-0.107
	(1.24)	(-2.83)
Two months ago returns	-0.083*	0.148*
	(-5.95)	(5.72)
Two months ago gains	0.064*	-0.082
	(3.14)	(-2.12)
Three months ago returns	-0.058*	0.087*
	(4.16)`	(3.60)
Three months ago gains	-0.093*	0.001
	(4.55)	(0.03)
Average returns 4-12 months	-0.010	0.550*
	(-0.48)	(13.04)
Average returns 13-24 months	0.134*	0.198*
	(6.12)	(4.61)
Average returns 25-36 months	0.080	0.055
	(4.06)	(1.32)
36-month standard deviation	0.003	-1.3 E-04
	(0.22)	(-0.01)
F-test for time fixed effects	33.33*	2.09

[a]New money represents additions or withdrawals. More money was withdrawn than added so the mean was negative (-0.83% per month).

Since performance persistence is small relative to the noise in the data, it is important to use lots of data. The four-year and three-year selection periods used here may be too small. The regression approach used here would allow using all the data when some funds have two years of data and others eight. But, you would not want to use data where you knew the CTA had made a major change in the trading system or a fund had switched advisors. There will always be some possible benefit from subjective analysis.

The main advantage of the regression approach does appear to be in using all the available data. The methods used here were not shown to be substantially better than a usual modified Sharpe ratio (which is the mean divided by the standard deviation). Publications which report CTA and fund returns should include a return/risk measure such as a Sharpe ratio.

Because of the low predictability of performance, it would be difficult to pick the single best fund or CTA. Therefore, it might be better to invest in a portfolio of CTAs. Picking CTAs based on returns in the most recent year may even be worse than a strategy of randomly picking a CTA.

Summary

The research sought to determine and explain the level of performance persistence in managed futures. Performance persistence could exist due to either differences in cost or differences in the skill of the manager. Our results favor skill as the explanation since returns were positively correlated with cost. A regression model was first estimated which included the average fund return as a regressor. The regression model indicated some statistically significant performance persistence. The performance persistence is small relative to the variation in the data (only 2-4% of the total variation). But, the performance persistence is large relative to the mean.

The regression method was expected to be the method most able to find performance persistence. Monte Carlo methods showed that the methods used in past research could often not reject false null hypotheses and would reject true null hypotheses too often.

Out-of-sample tests confirmed the regression results. There is some performance persistence, but it is small relative to the noise in the data. A return/risk measure showed more persistence than either of the return measures. There is some possibility of picking the best funds, but precise methods and long time periods are needed.

CTAs using short-term trading systems had lower returns than CTAs with longer trading horizons. CTAs with higher historical returns are now charging higher fees. CTA returns decreased over time and more recent funds have lower returns. At least part of this trend is likely selectivity bias. As dollars under management increased, CTA returns decreased. The findings of fund returns decreasing over time and as dollars invested increase suggests that funds exist to exploit inefficiencies.

The dynamics of returns showed small negative correlations for returns in the short run - especially for losses. The net effect over three years is positive which is consistent with a small amount of performance persistence. The withdrawal of dollars from CTAs shows that investors weight the most recent returns more than would be justified by changes in expected returns.

While several different methods were used, the results paint a consistent picture. To pick CTAs or funds based on past returns, several years of data are needed.

353

References

Elton, J., M. Gruber, and J. Rentzler. "Professionally Managed, Publicly Traded Commodity Funds." *Journal of Business* 60(1987):175-199.

Goetzmann, William N., Jonathan Ingersoll Jr., and Stephen A. Ross. "High Water Marks." Working paper, Yale School of Management, July 9, 1997, (http://viking.som.yale).

Irwin, Scott H., and B. Wade Brorsen. "Public Futures Funds." *Journal of Futures Markets* 5(Summer 1985):149-172.

Irwin, S., T. Krukemeyer, and C. Zulauf. "Are Public Commodity Pools a Good Investment?" *Managed Futures: Performance Evaluation and Analysis of Commodity Funds, Pools, and Accounts*. Chicago: Probus Publishing Company, 1992, pp. 403-434.

McCarthy, David F. "Consistency of Relative Commodity Trading Advisor Performance." University College Dublin, Dublin, unpub. thesis, 1995.

Schwager, Jack D. *Managed Trading: The Myths and Truths*. New York: John Wiley & Sons, 1996.