RCM Internal Paper 2012-05: Alternative Metrics for Portfolio Construction

January 22, 2013

1 Introduction

Diversification has been described as the only true "free lunch" for an investor. By simultaneously investing in a number of disparate and unrelated assets and/or strategies, a mathematical inevitability arises wherein returns can be maintained while intermittent losses (aka "drawdowns") are reduced. This is, in fact, one of the central tenets of Markowitz's modern portfolio theory (MPT). By using mean-variance approaches to optimize portfolio weights (the "mean" being the average return while the "variance" is essentially synonymous with the drawdowns), one eventually arises at what is known as the "efficient frontier". On the efficient frontier, the portfolio is optimized such that there is a maximal expected return for a given amount of risk. Equivalently, the approach minimizes the risk for a given expected mean return.

Critics have noted a variety of shortcomings of MPT when applied to "real-world" scenarios. We focus on two specific critiques: first, MPT considers only the first and second moments of the return distribution, i.e. mean and variance. It does not consider higher-order moments (specifically, skewness and kurtosis) even though these are arguably as important in a real-world setting. Second, the efficient frontier is only relevant to the extent that inter-strategy correlations are constant. Our experience suggests that, in the futures space, the constant-correlation assumption is often not valid, and to complicate matters, the extent of its invalidity depends on sometimes-subtle structural connections between strategies. Adding to the complexity is the fact that many managed-futures programs are actually multi-strategy in nature; this begs the question as to whether one can even arrive at true inter-*strategy* correlation measures. We are not aware of any studies that address these issues in detail. Yet in our experience, these are the dominant factors that causes portfolios to eventually exhibit sub-optimal risk/reward tradeoffs.

In a two-part series of non-technical papers [1, 2], we first introduced the challenges and benefits of adding short-term trading strategies to a portfolio of trend-following models. We extolled the benefits of diversification but noted that capacity constraints cause short-term systems to be fairly rare. In the second paper, we explored some general guidelines for moving beyond mean-variance, efficient-frontier-esque optimization methods. Starting with a trend-following core, we opined that there are three primary desires for the candidate programs that are needed to "round out" the portfolio. The first is solid risk-adjusted returns, the second is a low correlation to trend-following, and the third is a positive skew of returns (i.e. no "fat tail" of downside risk). We then pointed out that internal research has shown that any two of these properties is fairly achievable, but satisfying all three simultaneously is very difficult. This heuristic argument was made without the benefit of our new familiarity with higher-frequency (intra-day) trading strategies; we posit that such models can cut the Gordian knot that we believe exists for end-of-day strategies.

In this paper, we attempt to expand these concepts and subsequently utilize this knowledge to improve upon the standard methods of portfolio construction. We first examine the effects of higher-order moments on portfolio fitness; specifically, we consider their mutability for a given trading style and also the relevance of such moments on various time scales. Next, we investigate the number of **truly**-independent return streams from which the portfolio designer can choose. To assess this value, we first introduce the idea of "idiosyncratic noise", which is the random variability among similar strategies that disguises itself as diversification and can hence lead the portfolio designer astray. Armed with these ideas, we next analyze a number of trend-following and short-term systems using various correlation metrics. This analysis provides confirmation of the finite return streams available to the designer, and it also highlights the simultaneous challenges and benefits of adding short-term systems to a portfolio. Finally, we introduce a completely-novel method for portfolio construction, which is to categorize systems based on stressor- or shock-response analysis. Although it completely ignores statistical measures of the candidate systems, we will see that it can naturally lead to a robust, diversified portfolio that exhibits statistical "goodness" as an end-product of the optimization process.

2 Existing research

Modern portfolio theory dates back to the seminal work of Markowitz over half a century ago [3]. Since then, the concept has been extensively studied and extended and is now standard fare for many finance textbooks [4]. In terms of portfolios specifically incorporating futures-based strategies (generally trend following), there are many studies that have examined the symbiotic relationship of futures and stocks. It has been shown empirically that trend following tends to provide very good downside protection during bear markets and is in fact historically profitable during such periods [5, 6] (this should be contrasted with other bear-market hedges such as short-selling and out-of-the-money option buying, both of which tend to be unprofitable in the long-term and hence should properly be considered insurance and not true investments). Kaminski [7] generalized this argument in a recent study that shows futures-based strategies to be a source of positive returns in times of financial, economic, and geo-political crises, thus providing what she coined "crisis alpha".

Due to these inherent, stable structural differences between traditional "long-only" investments such as stocks and active investments such as managed futures, mean-variance optimizations and the subsequent efficient frontier calculations repeatedly show great benefit in adding managed futures to a traditional portfolio [5, 8]. Rollinger extends this analysis by looking at the first four moments (mean, variance, skewness, and kurtosis) of a portfolio consisting of varying percentages of stocks, bonds, equity-centric hedge funds, and futures. He found that, given a baseline stock/bond portfolio, futures strategies were as effective as hedge funds in improving the Sharpe ratio (mean/variance) but much *better* in providing neutral to positive skew and low kurtosis. In

Trend followers	Short-term traders
MAN AHL	Boronia
Aspect	Conquest (Macro)
Campbell	Crabel (Multi-product)
Cantab	Ion
FX Concepts (Multi-Strategy)	Kaiser
Lynx	QIM
Transtrend	Niederhoffer (Diversified)
Winton	Revolution (Mosaic Institutional)
	Revolution (Alpha)

Table 1: List of programs used in this study. All trend followers are in the Newedge CTA Index. All short-term traders are in the Newedge Short-term Traders Index except for the Alpha program.

fact, the optimal mix was found to be roughly 50% stocks/bonds and 50% futures.

While Rollinger's work certainly incorporates some of the same metrics we deem important to proper portfolio construction (e.g. higher-order moments of the return distribution), our goal is to augment these with more-radical concepts such as conditional correlations, time-scale-dependent statistics, and shock response analysis. In the first of our previous two-part series [1], we limited our discussion to the diversification benefits of short-term trading systems, along with the associated risk management and capacity challenges. We relegated a discussion of portfolio-construction concepts to the second paper [2], but in this treatise we provided only general guidelines for implementing alternative portfolio optimization methods. This paper augments the previous work by delving into much greater detail, providing a systematic approach and well-defined metrics to gauge the goodness of the final portfolio.

Despite years of dedicated research addressing portfolio diversification (why, how, how much), we still have more questions than answers. This paper should be viewed for what it is: a snapshot of an ongoing effort to better understand the full scope and potential of systematic futures-based trading strategies. Our hope is that every reader can find at least one valuable kernel of information, or perhaps at least a perspective not previously considered, that will help in their own efforts to generate "optimal" portfolios.

3 Data sets

We will make use of several data sets in this paper. First, we use the Newedge CTA Index as a benchmark for a "standard" diversified futures portfolio. The index is comprised largely of trend-followers but includes currency strategies and short-term systems as well. Most multi-manager funds exhibit a high correlation to this index, so it is a reasonable proxy for multi-strategy, futures-based portfolios. Next, we make use of the monthly returns of several trend followers included in the Newedge CTA Index. To augment these CTAs with a variety of short-term traders, we also include most of the constituents in the Newedge Short-term Traders Index. Finally, we also include Revolution's Alpha program with the other short-term traders since it meets the criteria

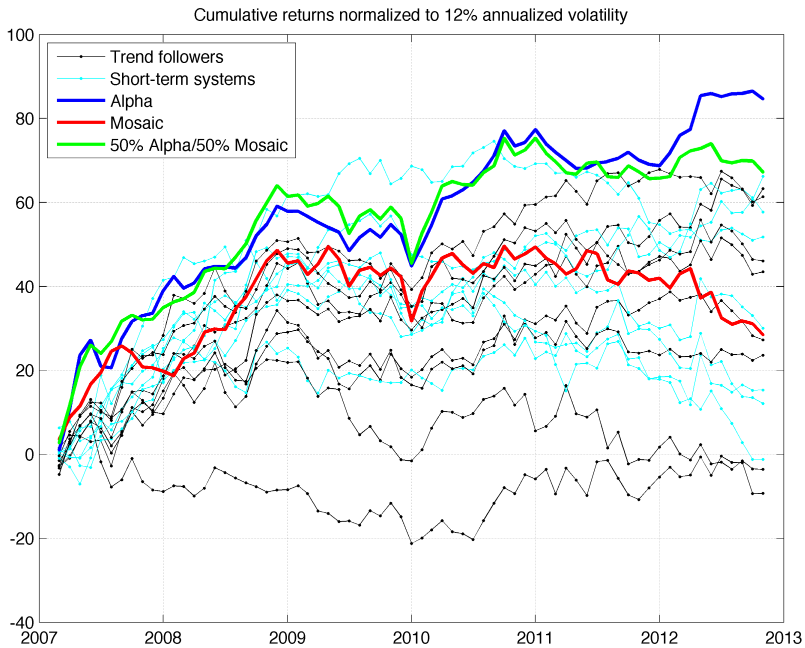

Figure 1: Cumulative returns of programs since March 2007.

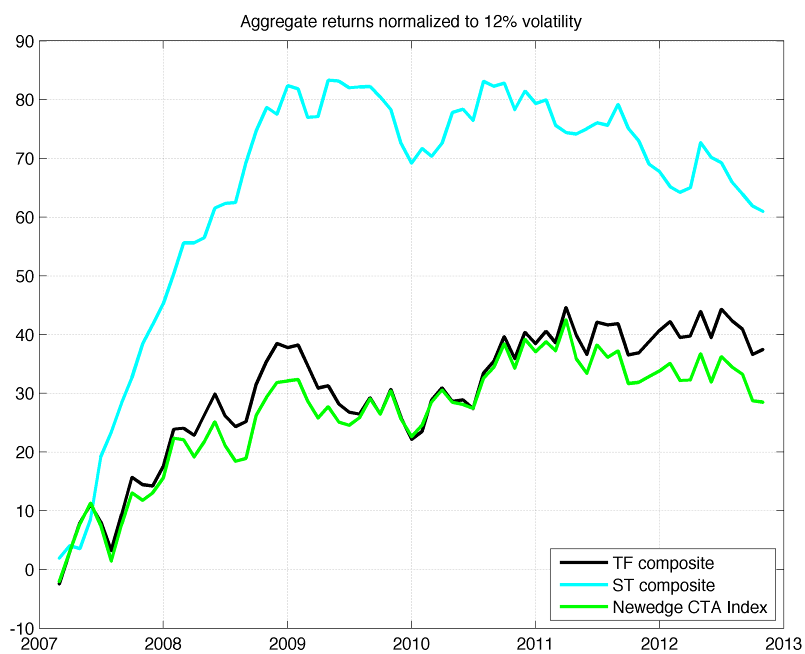

Figure 2: Average cumulative returns since March 2007.

and because there are structural differences between Mosaic and Alpha that are germane to our findings. Table 1 shows the full set of 17 programs used in this study. We have only monthly returns available for the non-Revolution programs; we use the time period from March 2007 to November 2012 (inclusive) in order to allow a full data set for all programs (we have Cantab data starting in March 2007; Ion started in November 2007 but we have back-tested simulation data that they published for January 1997 through October 2007). We do not attempt to adjust time-dependent volatility *within* each data set (for example, Winton's target volatility appears to have dropped noticeably over time), as this is too speculative. We do, however, normalize each overall set of returns to a 12% annualized volatility when doing comparisons.

Figure 1 shows cumulative returns of all the systems between March 2007 and November 2012. All systems were normalized to a 12% annualized volatility (based on monthly returns) for ease of comparison. The Alpha and Mosaic (Institutional) programs are shown in blue and red, respectively. To gauge our overall performance, we show a 50/50 mix of Alpha and Mosaic in green. Taken collectively, our performance has been quite respectable. However, since 2010 the counter-trend systems in Mosaic have struggled, with 2012 showing the greatest divergence between Alpha and Mosaic performance. Over the 69-month period, Mosaic is right about average relative to the other short-term systems, while Alpha is at the top of the performance comparison for this group of programs.

Figure 2 shows the average cumulative return for both the short-term and trend-following groups, again normalized to 12% annualized volatility. For comparison, the Newedge CTA index (adjusted to 12% volatility) is shown as well. The TF curve tracks this index extremely closely over the entire period. Conversely, short-term systems as a whole heavily outperformed during 2007 and 2008, treaded water between 2009 and 2010, and since then have been steadily unprofitable for the past two years. In Sections 5 through 7, we take a closer look at the possible reasons for this behavior and also at the ongoing evolution of our systems to counteract these recent challenges.

4 Skewness and kurtosis considerations

Skewness and kurtosis are familiar quantities to most portfolio builders, but we believe that there is still novel information to be gleaned from these relatively-pedestrian metrics. As a quick refresher, skewness refers to the asymmetry of a distribution (with positive skewness denoting a tendency for more positive outliers and a median that is less than the mean), while kurtosis refers to the likelihood of extreme events and is usually measured relative to a Gaussian (aka "normal") distribution. A Gaussian distribution has a kurtosis equal to 3, so "excess kurtosis" is simply kurtosis minus 3; positive excess kurtosis implies a higher likelihood of outlier events than would be experienced with a Gaussian distribution.

With investment strategies, positive skew is often desirable as it implies some increased likelihood of large gains without a corresponding likelihood of large losses. Kurtosis is a double-edged sword since a positive value generally implies a higher likelihood of both gains *and* losses. While most analytics software computes global mean, variance, skewness, kurtosis, etc. for the entire return distribution on some time scale (which is usually daily or monthly depending on the availability of data), what is not often employed is the use of *conditional* statistics. We have found two

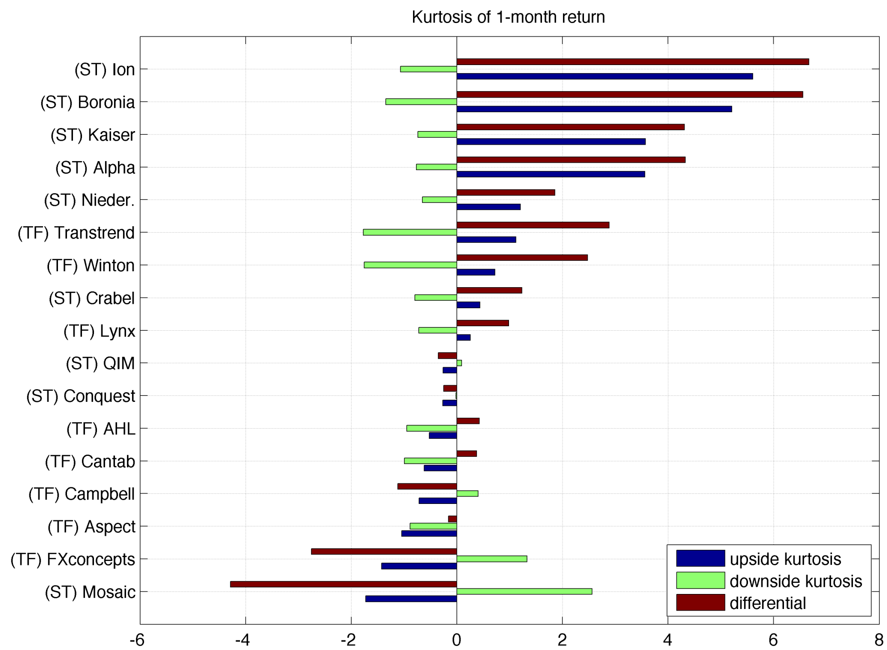

Figure 3: Excess kurtosis values of monthly returns. We show upside kurtosis, downside kurtosis, and the difference between the two.

very useful ways in which to first condition the data: first, constrain it to be over a *set* of time scales; second, constrain it to be either the set of positive *or* negative values, but not both. In this way, we can simply and easily decompose the global kurtosis value and examine the propensity for positive outliers versus that for negative outliers.

In addition, many short-term strategies are time-scale-dependent. Those employed by Revolution, for example, have positive excess downside kurtosis over time scales of 10 days or less, but this kurtosis largely disappears on time scales of 20 days or more. And rather than being a statistical artifact, we can explain this phenomenon due to the structural characteristics of the manner in which the systems accumulate and shed risk [9]. Although a portfolio builder may not have access to the inner workings of a particular system, a clever analysis of the output can provide valuable information regarding the system's structure.

Figure 3 shows upside and downside excess kurtosis for monthly returns; the programs are sorted in descending order of excess upside kurtosis. We separate the measurements between positive and negative returns in order to isolate specific features of each program's opportunity and risk. First, we note that most programs have negative downside kurtosis. In other words, one-month losses appear to be bounded more than one would expect with Gaussian data. Mosaic, however, is the most notable exception to this (which we expect based on its trading style). Second, on average, there is also a tendency for positive upside kurtosis. This is especially pronounced with several of the short-term programs but less so with the trend followers. Alpha is a notable example of this. Finally, the short-term traders (ST) have a larger range of outcomes than do the trend-

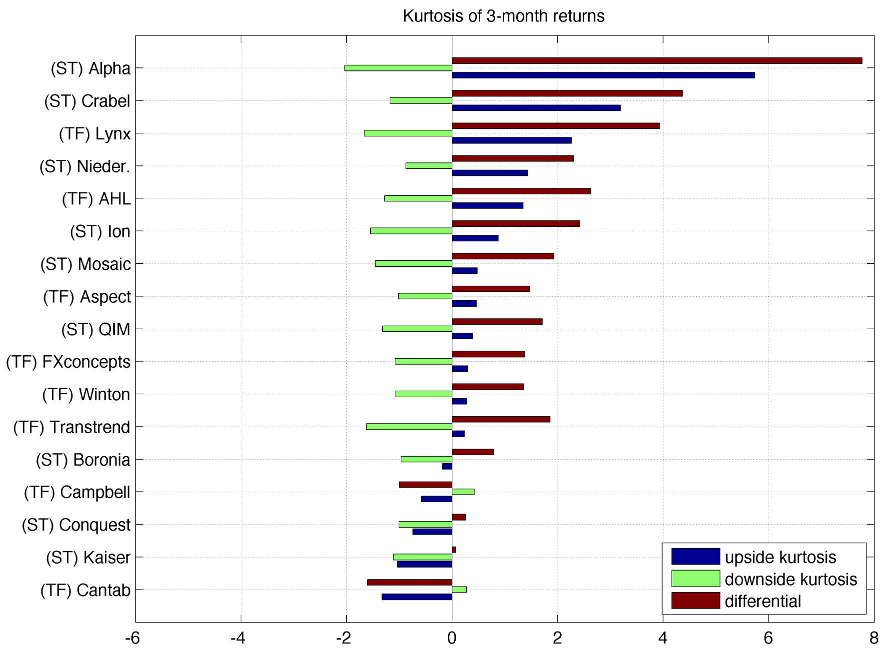

Figure 4: Excess kurtosis values of 3-month running returns. We show upside kurtosis, downside kurtosis, and the difference between the two.

followers (TF). This is perhaps unsurprising given the wider gamut of strategies employed by the short-term group.

Figure 4 shows the excess kurtosis values for 3-month running returns. The picture is now substantially different. Boronia and Kaiser (near the top of 1-month graph) are now near the bottom, and Mosaic (at the bottom of the 1-month graph) is positioned in the middle of the pack with positive upside and negative downside kurtosis. Alpha has moved to the top of the chart. Between the TF and ST designations, there is little difference on average. There is admittedly much random noise in these measurements. The 99% confidence interval for both figures is about +/- 1.2, so values close to zero are insignificantly biased. But the overall results reveal what we have identified as inherent structural differences across strategies. We have done considerable research exploring how these structural differences give rise to necessarily-different return profiles [9]. The details of the studies won't be repeated here, but the overall conclusions are as follows:

1. The short-term distribution of returns depends critically on how the system accumulates and sheds risk. Increasing exposure in a profitable position (and vice versa) gives rise to positive upside kurtosis and negative downside kurtosis, even in random markets. This suggests that many of the short-term CTAs are employing a momentum-based strategy on short time scales. In contrast, Mosaic and Alpha both accumulate risk, to a point, during unprofitable positions. Mosaic does this to a greater extent than Alpha and presumably more than the other ST systems as well, which explains its position near the bottom of the 1-month chart. However, both systems are also designed to reduce kurtosis as quickly as possible. This again is consistent with the 3-month rankings, which show both systems moving up substantially relative to the peer group.

2. Many of the trend-followers also likely employ some exposure to counter-trend models. In addition, we believe there may be deliberate intra-month deleveraging taking place when month-to-date returns are especially positive or negative. This results in smaller kurtosis values than might be expected from pure momentum traders. The longer trade holding periods (relative to short-term momentum players) also provide more opportunities for averaging away short-term price bursts while still in a given position.

3. Overall, there seems to be an asymptotic approach toward Gaussian behavior for all systems shown in this study. This should be contrasted with pure mean-reversion systems (aka "short volatility") such as option writing, which tend to show positive downside kurtosis even on long time scales due to small probabilities of very large losses. Revolution's systems show positive downside kurtosis on 1-month time scales or less, but the systems are designed to eventually reduce risk in the face of extended losses. This provides a normalizing mechanism on longer time scales.

In summary, short-term drawdowns and the consequent return distributions should be viewed as a function of both risk management **and** the style of trading, not just the former. This is because the two are inextricably linked. It is critical not just to measure and compare monthly kurtosis and skew values, but to understand *why* certain systems have significantly different properties on certain time scales. Mean-variance optimization approaches have no ability to address these points despite their critical impact on portfolio design.

Ultimately, the preference of one type of short-term system over another should depend to a large extent on the goals of the portfolio designer and the other constituents of the portfolio itself (i.e. finding the best complementary systems). To that end, we now examine methods to quantify the variety of complementary systems available to the portfolio designer.

5 Idiosyncratic noise, degrees of freedom, and the illusion of diversification

In an ideal world, a portfolio designer need only look long enough and hard enough, and eventually he or she will uncover a sufficient number of uncorrelated strategies with which to construct a "bulletproof" portfolio that generates returns with clock-like consistency. Reality, however, tends to thwart this utopian ideal. First, there is an issue of style. Most portfolio designers are mandated to work *within* a particular asset class, e.g. equities or futures, not *across* asset classes. This necessarily narrows the universe of possible strategies available to the designer. This is likely less of a constraint in the futures space, where all profitable strategies need to be actively managed (as opposed to passive indices) and hence more variety can be found. So if we constrain our search to the world of managed futures, how many different (i.e. uncorrelated) strategies exist? Foreshadowing the discussion of Section 6, this presumes that we have a preferred time scale over which to measure correlation, and it also presumes correlation values to be constant. For simplicity, let's assume the latter, and let's also assume that monthly returns are appropriate for determining correlation.

Before continuing, we need to introduce the concept of "idiosyncratic noise". We define idiosyncratic noise as the return variability that arises between managers not due to true differences in trading strategy but rather due to differences in market or sector weightings, volatility estimation approaches, trade entry/exit nuances, and any other particulars of the system **implementation** (as opposed to the system **concept**). Understanding idiosyncratic noise is critical, because on short time scales it can dominate a return stream and give the appearance of differentiation between programs when in fact none exists. That's not to say that focusing on idiosyncratic noise is a completely futile exercise; often times, even if a portfolio consists solely of one main strategy (e.g. trend following), the inclusion of multiple managers allows the idiosyncratic noise to average itself out and thus provides a way in which to generate a slightly smoother return over time (i.e. remove excess kurtosis). But averaging out such noise, by definition, provides no true diversification. It simply helps one extract the "alpha" of the underlying strategy more effectively.

The goal of this section is to strip away the idiosyncratic noise as much as possible and to attempt to measure the number of truly-independent strategies available to the portfolio designer. This is the raw material the designer needs in order to provide value. To that end, we present three approaches to estimating the number of uncorrelated strategies (aka the number of "degrees of freedom", or DOF) available for portfolio diversification.

5.1 Time-scale cascade

The first method to approximate the total DOF rests on a single, empirical observation: based on previous research, we typically see our strategies become uncorrelated when we modify the time scales by a factor of 4.[1] In other words, we change nothing in terms of the trading style or trade entry/exit criteria; we only change the lookback times over which we analyze the data to determine new trading signals. In directional trading, changes in price are required in order to generate a profit; simply put, we need a trend in prices on the time scale of our holding period. Armed with this viewpoint, we can simplistically yet reasonably conclude that **every** system is trend-following or at least trend-exploiting, further implying that the only true diversifier is time scale (we will present more evidence for this theory in Section 7). Starting with long-term trend followers, we estimate that their longest time scale is roughly 100 days and hence define this as DOF #1. Scaling this down by a factor of 4, the 25-day time scale is the next degree of freedom, DOF #2. DOF #3 is then roughly on 6-day time scales, while DOF #4 is on 1.5-day time scales. The fifth DOF is on a time scale of about 0.4 days, and this is where we stop counting. The reason is one of practicality. Our own highest-frequency system has holding periods of about 0.5 days or so, and we recognize that such a strategy has significant but still-limited capacity. Moreover, competitive analysis of other intra-day trading systems suggests that they operate on similar time scales or even a bit longer. A portfolio designer needing to accommodate limited capacity could possibly find some diversification on time scales of 0.1 days or less, but it is unlikely to be scalable. Even 0.4-day strategies could not be easily deployed to every multi-strategy, futures-based fund in existence; there simply isn't sufficient liquidity at such time scales. As a result, based on nothing but an empirical rule ("a factor of 4 provides de-correlation"), a trend-following time scale of 100 days, and a practical limit on the capacity of high-frequency systems, we arrive at an estimated 4-5 DOF available to the portfolio designer. This analysis ignores any and all nuances of trading system design and simply assumes one dominant, exploitable pattern per time scale. This isn't necessarily unreasonable since prices can't be both **globally** trendy and mean-reverting at the same time (though strategies that are conditioned to separate out trendy from mean-reverting periods could in theory produce two different strategies per time scale). With this possibility in mind, the next two sections will attempt to corroborate this estimate with more-refined and more-direct techniques.

5.2 Correlation-based estimation

We now utilize correlation measurements to facilitate the second estimation method of the total DOF. Given the somewhat unorthodox techniques presented in this section as well as in Section 5.3, Appendix A contains a proof of concept of these approaches using artificial, randomly-generated data. Sections 5.2 and 5.3 should, however, be read first in order to fully understand the methodologies utilized.

Figure 5 shows the correlation matrix of 1-month returns. The programs are sorted such that

[1] Intuitively, one might expect that a factor of 2 would be sufficient. For sinusoidal functions with integer wavelength constraints, any wavelength ratio of 2 or greater provides a correlation of 0. We have not delved into the reasons for the observed factor of 4 but simply use the result as-is.

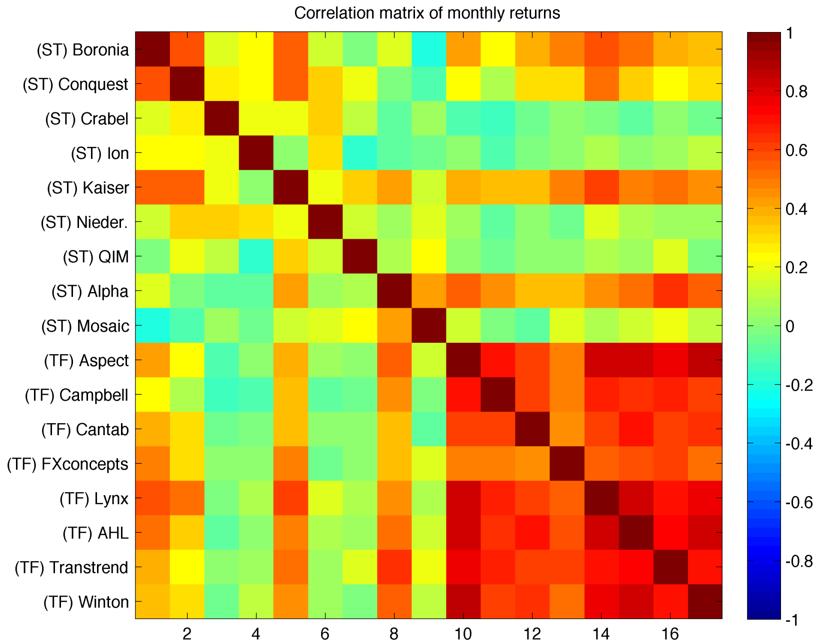

Figure 5: Correlation matrix of monthly returns.

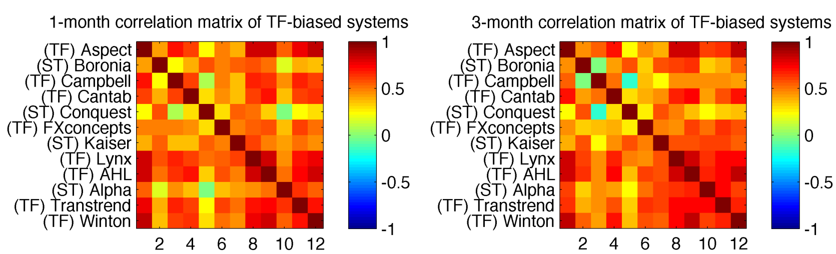

Figure 6: Correlation matrices of TF and short-term, trend-biased systems for 1-month returns and running 3-month returns.

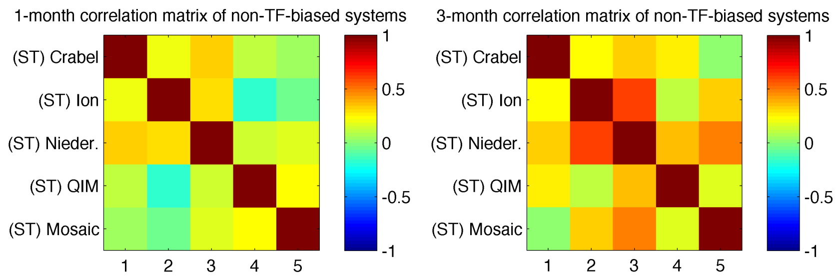

Figure 7: Correlation matrices of non-trend-biased ST systems for 1-month returns and running 3-month returns.

short-term (ST) systems are at the top and trend-following (TF) systems are on the bottom of the figure. A few observations are clear: first, all TF systems have a high correlation to each other. Second, several of the so-called ST systems have a substantial correlation to trend-following as well. These are Boronia, Kaiser, Conquest (to a slightly lesser extent) and our own Alpha program. The inclusion of Alpha is not at all surprising since it is designed to have a 0.5 long-term correlation to the Newedge CTA index. However, it does indicate that at least 2-3 of the Newedge Short-Term Traders Index constituents are measurably aligned with trend-following returns; we will refer to these four ST systems as "trend-biased" for the remainder of this discussion. The TF systems and the trend-biased ST systems are shown in their own matrix correlation graph in Figure 6. The non-trend-biased ST systems are shown in Figure 7. It is clear that these systems are much less similar to each other, though the average correlation does grow on longer time scales and thus indicates some of the apparent diversification is likely from idiosyncratic noise.

The set of programs is thus somewhat tri-modal. The TF programs are extremely similar to each other, while a subset of the ST programs have a noticeable but incomplete overlap with the TF systems. The remainder of the ST programs are distinct from trend-following and are also largely distinct from each other (or at least appear to be) based on monthly returns. The use of the correlation matrix shown in Figure 5 is thus a bit tricky from which to extract a DOF estimate. As one adds trend-following models, these dominate the average correlation value. However, if we separate out TF and trend-biased ST programs from the other ST programs, it is difficult to ascertain the true overall DOF because the non-trend-biased ST ensemble is still not completely

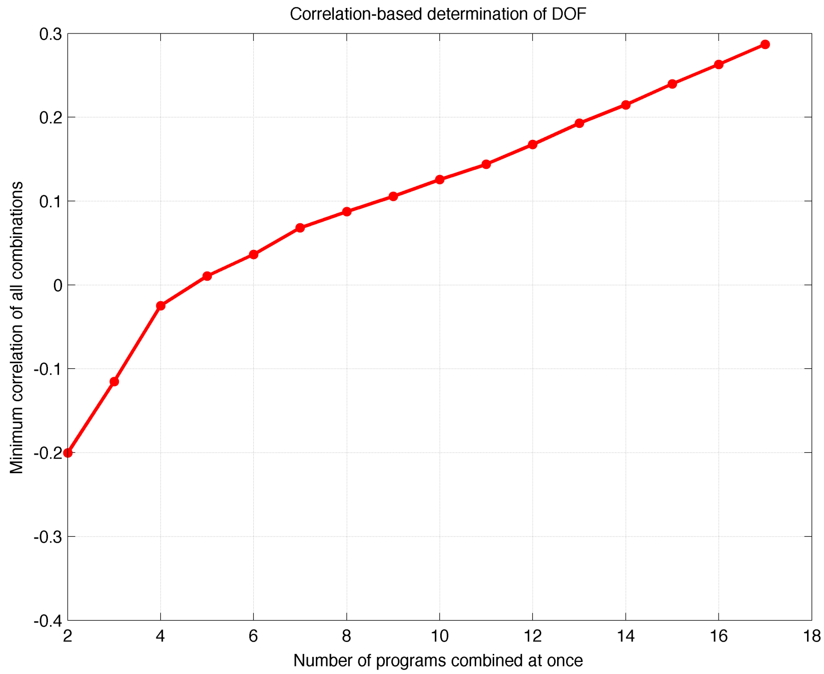

Figure 8: Minimum correlation obtained by taking N programs at once, where N varies from 2 to 17.

independent from the trend-followers (i.e. there is a small but positive correlation between the two). Moreover, there is also incomplete overlap between the TF programs and trend-biased ST programs, so treating these as one strategy seems inaccurate (we will confirm the distinction between the two in Section 5.3).

Our proposed solution is as follows: we consider all M combinations of N total strategies, where N is incremented from 2 to 17 (i.e. the total number of programs in this study). The value of M is given by the combinatorial rule

$$M = \frac{17!}{N!(17-N)!}, \tag{1}$$

where ! is the factorial function. For any value of N, if there are really N uncorrelated strategies and hence N true degrees of freedom, then some combination(s) among the M choices should give us a correlation of 0 or less. In other words, given a basket of 17 systems from which to pick, we might decide to look at all combinations of five managers. Many of these combinations would just end up consisting of five trend followers, but some combinations might reveal a more-desirable mix of five uncorrelated strategies. Given the random noise in the returns, we could also choose to take the correlation at the 5th percentile or something similar in order to have more confidence that the value is truly less than 0, but this likely wouldn't change the results much.

Not knowing the number of truly-uncorrelated strategies *a priori*, we simply increment N until the minimum measured correlation becomes positive. Figure 8 shows the minimum correlation obtained versus the number of models N considered at one time. Consistent with the previous findings, we see that the minimum correlation becomes positive between N=4 and N=5. This suggests that these 17 strategies have only 4-5 degrees of freedom between them, since for values larger than N=5 we can't find any combination where the strategies have an average pairwise correlation of 0 or less (meaning that every additional strategy combined into the best 5 overlaps in some way with them). Since the trend-following models are highly-representative of their niche, and since the short-term models have such low monthly co-correlations, it seems likely that these 17 programs fairly represent the universe of possible systems (at least in kind if not in relative frequency of occurrence). For rigor, this exercise could be repeated with a much larger number of programs and a more rigorous statistical bound on the correlation values in order to offset random outliers; we have not performed this exercise but expect that it would produce similar results. This method is fairly simple but potentially quite useful, as we can do the analysis on a number of desired time scales and examine the number of independent return streams to expect on a daily, weekly, monthly, quarterly, or yearly basis.

Unlike the time-scale cascade exercise, this approach uses real-world data from currently-traded systems. The primary shortcoming of this approach is that we may not have sampled the space of possible strategies completely since the onus is on us to include a sufficient mix of programs with which to start. Common-sense arguments, however, dictate that any marginally-obvious strategies would have long ago been discovered and put into practice. We now turn to our final DOF estimation method, principal component analysis, to see if this estimate is consistent with the expectation that there are 4 to 5 degrees of freedom available to the portfolio designer.

5.3 Principal component analysis

Principal component analysis (PCA, also called singular value decomposition or empirical orthogonal function analysis) can be highly useful in extracting information from high-dimensional data sets, but the results require careful interpretation. The methodology seeks to find a set of orthogonal components that most **efficiently** describe the input data, and it does so by maximizing the amount of variance that is explained for any given number of orthogonal components. The upshot is that the method allows us to explain as much of the variability in the data as possible (which in our case are the return streams) with as few independent sources as possible (which are presumably the true degrees of freedom).

In our case, we have 17 time series of returns. If all these happened to be identical, then PCA would show that one principal component could explain all the variance of the entire set of data (in other words, there is really only one strategy being pursued, and it is the same for all systems). Similarly, if all the data sets were identical except for a small amount of random noise, then the first principal component would consist of the "signal" (and hence explain the vast majority of the variance) and the remainder would represent the low-level random noise. In this case we would still be able to conclude that almost all of the variability could be explained by one principal component and that the input vectors are essentially identical, with the difference likely being the idiosyncratic noise previously introduced. Extending this concept slightly, we can imagine a

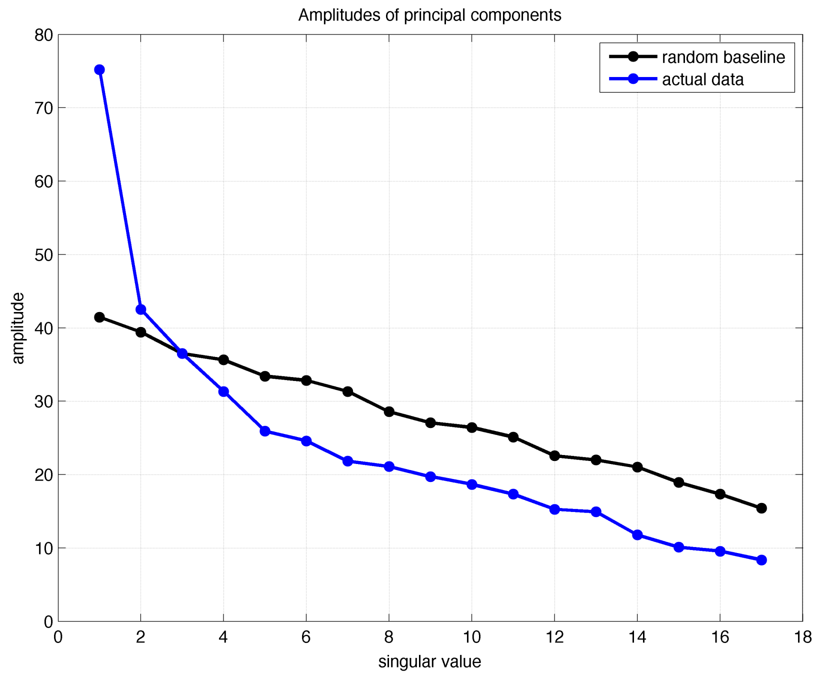

Figure 9: Principal component amplitudes for actual systems in blue and random baseline in black.

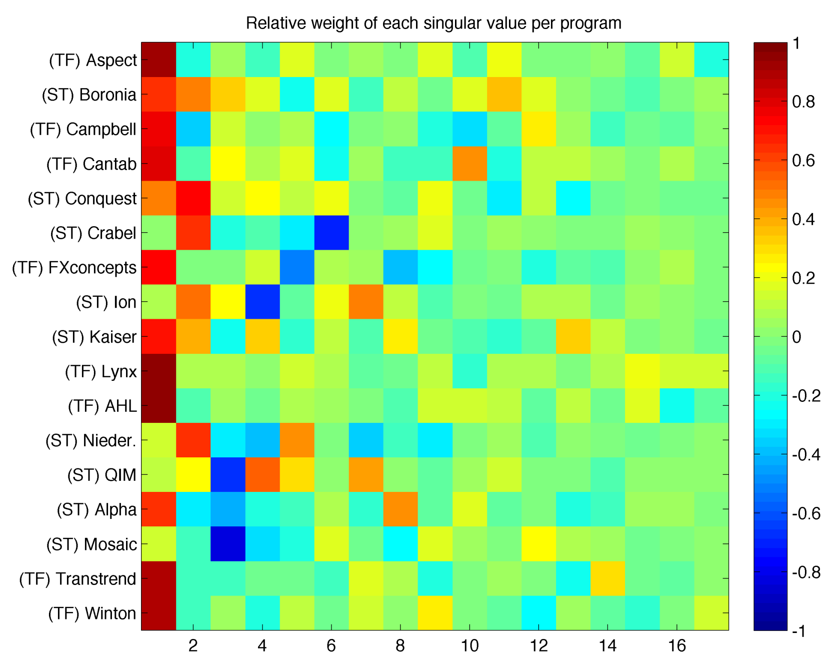

Figure 10: Amplitude of each principal component for each program. Negative values indicate that the system has a negative correlation relative to that component.

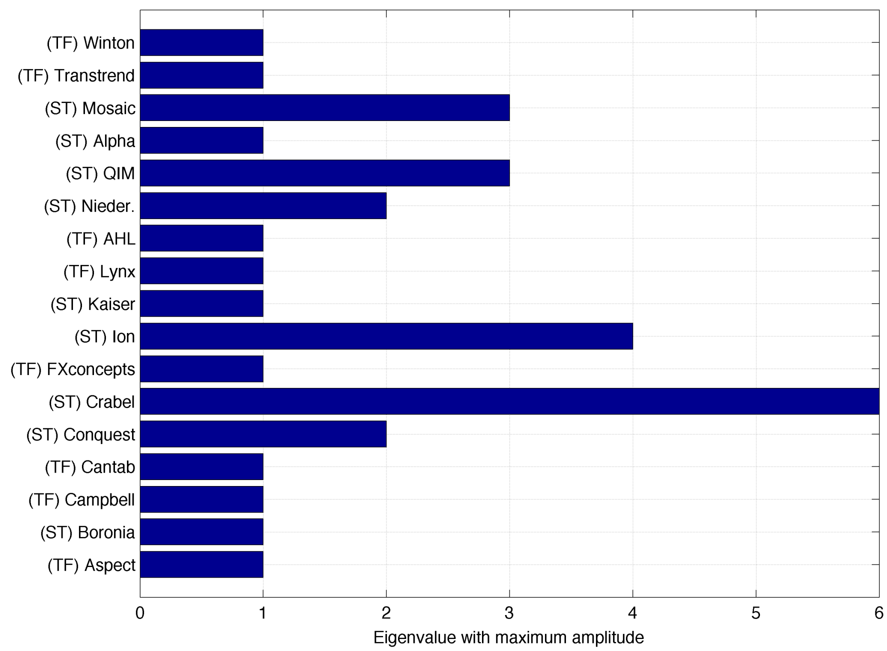

Figure 11: Principal component of maximum influence for each program. Note that all TF programs are dominated by principal component 1.

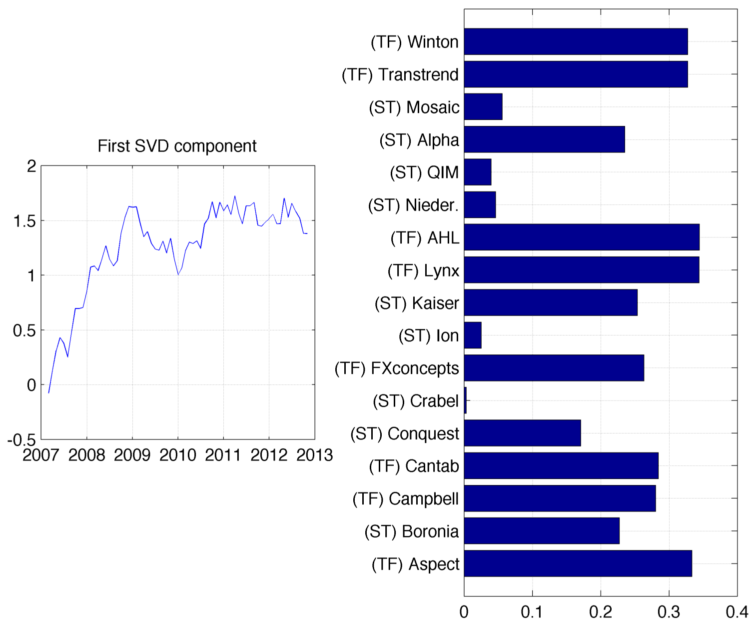

Figure 12: Time history of cumulative return of principal component 1 (with arbitrary mean added). The bar chart on right shows the influence of this component for each program.

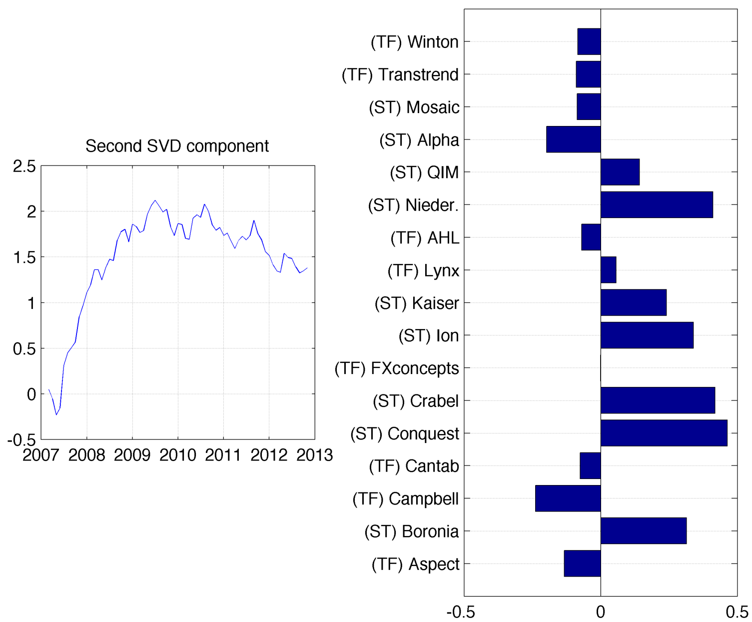

Figure 13: Time history of cumulative return of principal component 2 (with arbitrary mean added). The bar chart on right shows the influence of this component for each program.

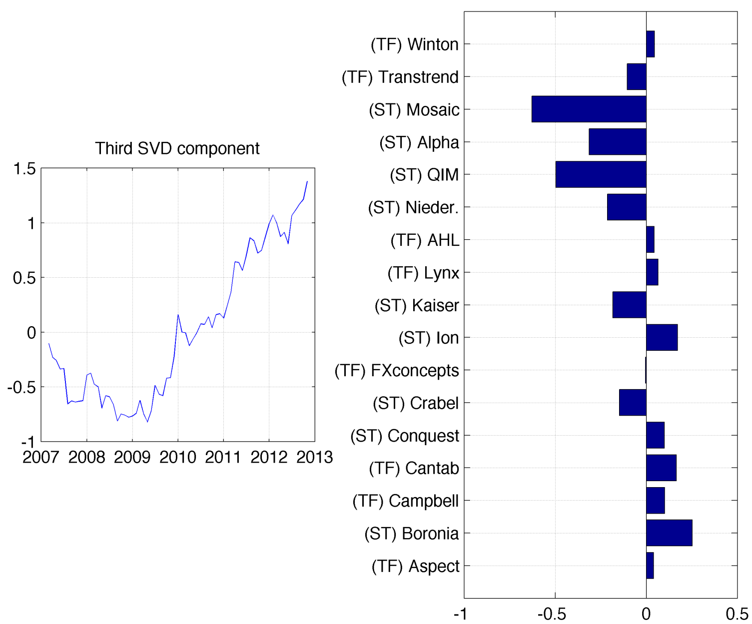

Figure 14: Time history of cumulative return of principal component 3 (with arbitrary mean added). The bar chart on right shows the influence of this component for each program.

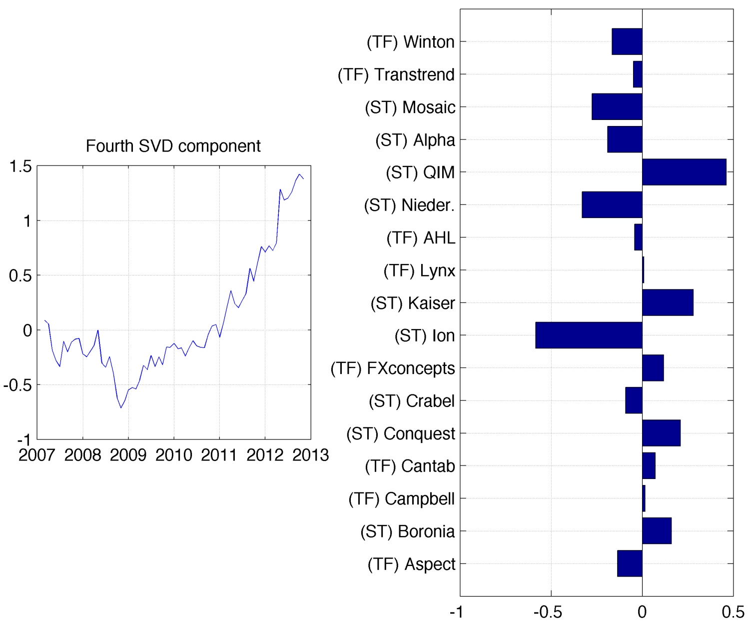

Figure 15: Time history of cumulative return of principal component 4 (with arbitrary mean added). The bar chart on right shows the influence of this component for each program.

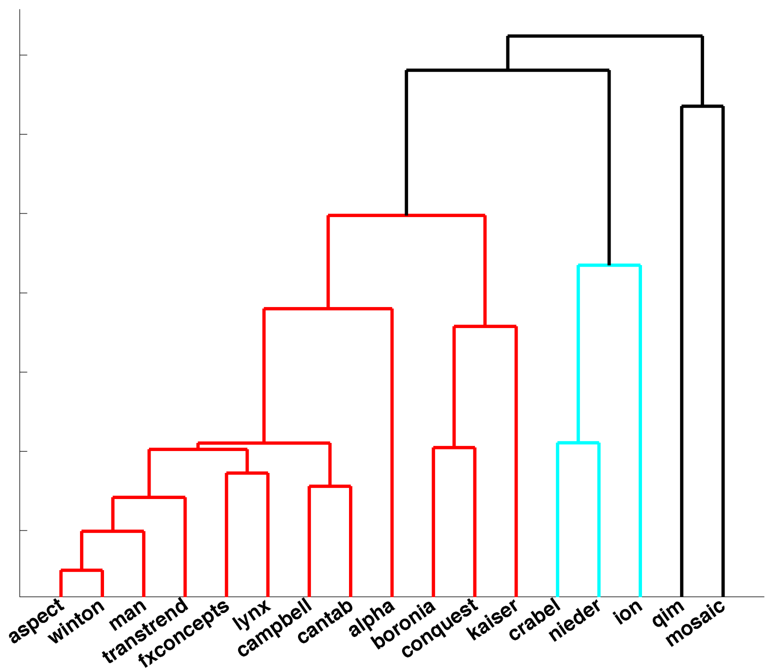

Figure 16: Output of `dendrogram()` classification routine in Matlab. Note that all TF programs are on the left side and all ST programs are on the right. Mosaic is most closely associated with QIM, while Alpha sits right in between the TF and ST groups.

situation where all input vectors are different linear combinations of two independent sources. This would occur, for example, if all 17 CTAs employed 1 long-term trend-following system and some other, independent short-term system. In this case, PCA would yield two principal components (the trend-following return stream and the short-term system's return stream) to explain the set of input vectors.

Although this sounds like a near-magical technique (and can be in many cases), there are two difficulties with the approach. First, since it is explaining the *variance* in the data set, the number of times a particular pattern occurs will influence the results because it affects the portion of total variance comprised by a single pattern. Therefore, frequency of representation matters in terms of principal component rankings. Second, because the principal components are simply numerical outcomes of variance maximization (equivalent to a coordinate rotation of the data), the principal components themselves can be difficult to interpret. There is no mandate that they need to make intuitive sense. A better approach may be to decompose the original data set using the concept of *independence* as opposed to *orthogonality*; this can be done via independent component analysis (ICA) [10], but we have not tried this technique. Whatever the deconstruction method, to make the results most useful, one needs to tie the principal components back to interpretable processes. In our case, this means that we'd like to be able to show that each principal component looks like some return stream we recognize and understand.

Figure 9 shows the amplitudes of each principal component, along with results for random data as a baseline. The first component clearly dominates, and we will see shortly that this represents trend following. Also note that, beyond 4 to 6 principal components, the amplitudes decay in line with random noise. This suggests that most of the components are not structural in nature but rather due to the idiosyncratic noise that dominates monthly returns.

Figure 10 shows the entire spectrum of amplitudes; each program is broken out to show the presence of each principal component in its return stream. Negative amplitudes simply mean that the program in question opposes the sign of the principal component (e.g. if component 1 represents trend following and someone exhibited a negative amplitude, it would indicate they have some component that universally trades opposite the trend). Figure 11 shows which principal component is dominant for each program. For all of the trend followers and a few of the short-term traders, component 1 is universally dominant. This is unsurprising since this component reflects a trend-following return stream. The short-term traders are a more diverse group. Component 2 (which we will argue is a much shorter-term trend-following component) dominates for Niederhoffer and Conquest, while Component 3 dominates for QIM and Mosaic (both of these programs have a negative value for component 3, meaning they oppose this pattern; such opposition to the pattern implies a counter-trend strategy of some sort). Component 4 dominates Ion's returns, while Component 6 dominates Crabel (though component 2 is nearly as dominant for Crabel).

Figures 12 through 15 show the principal components themselves as cumulative returns over time. Since means must be removed before doing the analysis, we added back a reasonable mean value to give the curves a positive bias over time. This helps show when each component profits or not. The bar charts in each figure show the amplitude associated with that principal component for each CTA's program. The first component clearly looks like long-term trend following. Moreover, the amplitudes coincide with expectations, with Transtrend, Winton, AHL et al. exhibiting a strong bias towards this return stream. It is also interesting that no one directly opposes it (i.e.

all amplitudes are positive). This is consistent with our research that shows trends to be the only non-random dynamic on longer time scales.

Moving to principal component number 2, we recognize this cumulative return pattern from internal research. It lines up very closely to short-term trend following. The bar chart supports this assertion, as the long-term trend followers show little exposure to this pattern. Conversely, Kaiser, Crabel, Conquest, and Boronia have strong positive exposure, which is consistent with our understanding of their base strategies (or at least a part of their strategy sets). Ion and Niederhoffer also have a strong positive overlap with component 2, though we haven't previously examined their systems for a short-term trend-following tendency.

Looking at Figure 11, we see that 13 of the 17 programs are primarily represented by the first two components, meaning that few programs implement strategies that explore other degrees of freedom. Principal component 3, shown in Figure 14, is positive for Boronia and Ion but strongly negative for Mosaic and QIM (and less strongly negative for Alpha). We interpret this as a short-term reversal breakout mechanism; our own systems employ trend reversion strategies (among others) and we believe QIM uses structurally-similar strategies (which would mean that Mosaic, Alpha, and QIM would oppose this dynamic and hence exhibit negative amplitudes), so this makes sense. Finally, Figure 15 shows Mosaic, Niederhoffer, and Ion exhibiting a negative exposure to component 4, while QIM and Kaiser (and Boronia and Conquest to a lesser extent) exhibit positive exposure. This dynamic is harder to interpret, but it likely shows a dichotomy between short-term traders on particular time scales. Based particularly on May 2012, which shows a highly positive return for those with positive exposure to this principal component, those with positive exposure must employ momentum on some short time scale, while Mosaic, Niederhoffer, and Ion employ mean reversion on the same time scale. Mosaic, Alpha and Niederhoffer are the only programs with substantial negative exposure to components 3 and 4, thus highlighting a structural difference between Mosaic and most of the other short-term traders. Note, however, that Mosaic and Niederhoffer differ heavily on component 2, thus repeating the theme of partial but incomplete overlap between the short-term strategies.

As is probably evident, once the principal components cease to be readily identifiable (e.g. component 1 as long-term trend following and component 2 as similar to our internal, short-term trend-following models), they become difficult to interpret. We can infer the most-likely concept in many cases, but the orthogonality concept at the core of PCA means that we may not be able to easily replicate the pure return stream with a simple model. Nonetheless, differences in each program's exposure to each of the principal components at least indicates that there *are* differences in the various strategies. Presumably, differences in the conceptualization and construction of the various short-term trading strategies give rise to principal components that are present in significant positive *and* negative exposures, and two systems that align with respect to one component (e.g. QIM and Mosaic with component 3) can be diametrically opposed on another (e.g. QIM and Mosaic with component 4). In terms of total degrees of freedom, if we couple the amplitude decay shown in Figure 9 with the maximum amplitude index shown in Figure 11, we can reasonably conclude that there is no significance to components 7-17 other than to reflect idiosyncratic noise. This suggests 6 DOF, and even component 6 is only potentially meaningful because it explains a large portion of Crabel's return variability.[2]

Finally, we use the per-program amplitudes of the first four principal components to group them according to "similarity". This is done using the `dendrogram()` function in Matlab, which essentially makes a "family tree" of programs. Figure 16 shows the results. This strikingly illustrates the output of the principal component analysis in a very compact manner:

1. All the TF programs are on the left side of the graph, while all the ST programs are on the right side.

2. Mosaic is most similar to QIM, which is expected based on principal component 3.

3. The trend-biased ST programs (Conquest, Boronia, and Kaiser), are grouped together.

4. Alpha sits right in the middle of the 17 programs, bridging the gap between TF and ST styles as would be expected based on its mandate to have a 0.5 correlation to trend-following.

5. Man (AHL), Winton, and Aspect are all clustered together, perhaps reflecting the common ancestry of the firms' principals (all originally at Man).

Although it doesn't convey any new information *per se*, we find the family-tree graphic to be a powerful method by which to compress the principal component output into an easy-to-interpret visual aid.

5.4 Summary of DOF determination

We have utilized 3 different approaches, ranging from the purely theoretical (time-scale cascade arguments) to purely empirical (correlation and PCA analyses). All three of the analyses, somewhat surprisingly, show that there are only between 4 and 6 degrees of freedom from which to choose when building a portfolio. Based on a weighted interpretation of the findings, we conclude that the most likely value is 5 DOF. This does not mean that 5 programs can adequately span the space of profitable trading strategies. First, most strategies do not target an exact time scale but rather aim for a range. Furthermore, exposure to some dynamics are planned as part of the program's "alpha"; other exposure is a by-product and may generate an unwanted and unprofitable exposure to other time scales. Finally, some programs may more optimally extract alpha from a particular time scale than another; there is still a quality argument to be made even if the number of alpha sources is limited.

The findings also confirm another commonly-held belief about short-term systems relative to trend followers: this is the assertion that short-term traders are more diverse because they have more potentially-profitable dynamics to target. The principal component analysis indeed shows that long-term trend-followers are almost completely represented by component 1. The remainder

[2] We have not yet understood this component; it has some curious intermittent overlap with component 4, perhaps suggesting that the strategy mix in Crabel's Multi-Product program is changing over time. If so, then this component doesn't really represent a new DOF but rather explains a time-dependent influence of a mix of the first 5 components.

of the significant components are populated by various short-term systems. Based on our time-scale cascade arguments in Section 5.1, this makes sense. Three of the five independent time scales represent holding periods of 6 days or less; although one could change the starting point at the long end of the temporal spectrum (e.g. we could choose 60 days as the longest time scale instead of 100 days), the geometric nature of the analysis ensures that most of the degrees of freedom will always exist for holding periods less than 10 days. The challenge to the portfolio designer is twofold: first, there is great value in adding such systems but the capacity is somewhat limited (i.e. potentially strong demand with necessarily-limited supply), and second, each program is generally a combination of various degrees of freedom, making it difficult to anticipate when each system should profit. Small differences in system construction can lead to large deviations in short-term performance. This gives rise to the illusion of infinite diversification, because pairwise correlations can continue to look low even though the same 5 DOF are simply being re-mixed in different proportions in each additional program that is considered. Once the inevitable idiosyncratic noise is added on top of this, pairwise correlations can approach zero if the full set of potential managers is mined carefully (which probably happens inadvertently as databases are searched for promising new candidates). The danger of the diversification illusion is profound, as the unwary portfolio designer will have a false sense of the robustness and diversification of the portfolio. Blind data mining will invariably result in a portfolio that overpromises and under-delivers once set into motion. Avoiding this trap is critical, and in the next section, we explore correlation in greater detail and examine methods to make program selection more robust.

6 Correlation considerations

Our approach to analyzing correlation philosophically follows the skewness and kurtosis discussion in Section 4. Specifically, conditioning the data to look at correlations over different time scales and only when the correlation target is either positive or negative (but not both) reveals much about the true nature of the system and averages away much of the idiosyncratic noise. That being said, over extremely-long time scales (6 months or greater), we believe that correlation analyses have reduced validity and that one has to begin to account for macro-environmental considerations; we defer this discussion to Section 7. Before proceeding further, we introduce an observation that conceptually connects the discussion in Section 5 with the general correlation results shown in this section. Following a previous paper [1], there is a direct relationship between correlation and degrees of freedom. For an arbitrary number of systems where each and every pair has a correlation C, the diversification D achievable from that system is given as

$$D \rightarrow \frac{1}{\sqrt{C}}, \tag{2}$$

so that the diversification approaches this ratio as the number of systems becomes large. In the upper limit of C=1, it follows that D=1 for **any** number of systems (and which therefore means that no additional diversification was achieved), and in the lower limit where C=0, the diversification D approaches infinity given enough systems.

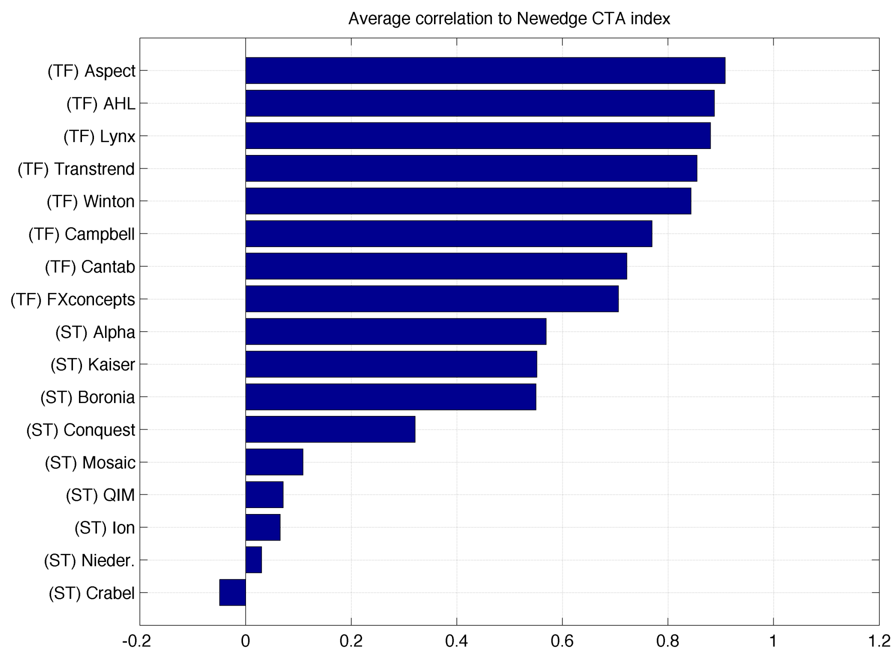

Figure 17: Correlation to Newedge CTA Index (March 2007 - November 2012).

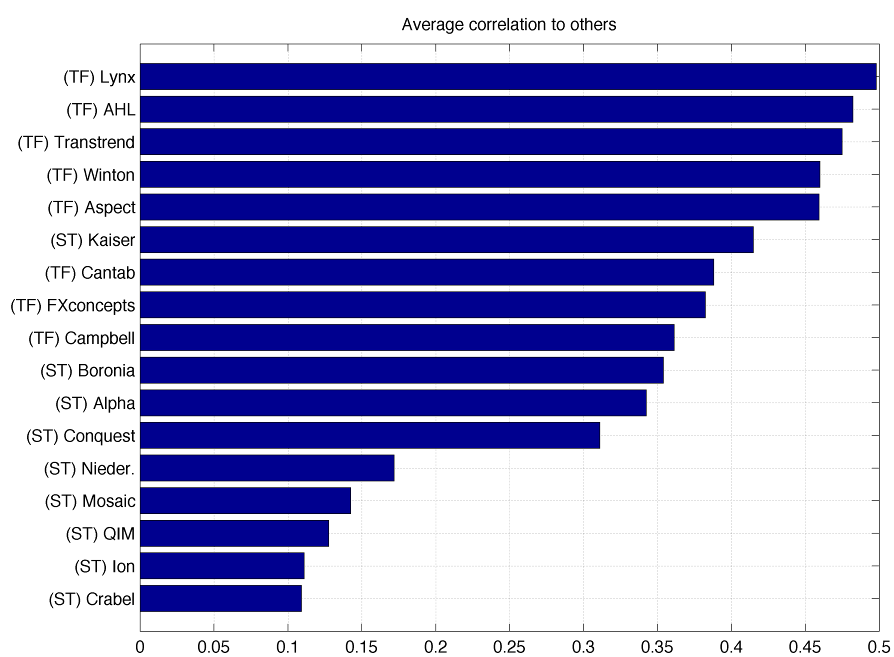

Figure 18: Average correlation of each program to the others (March 2007 - November 2012).

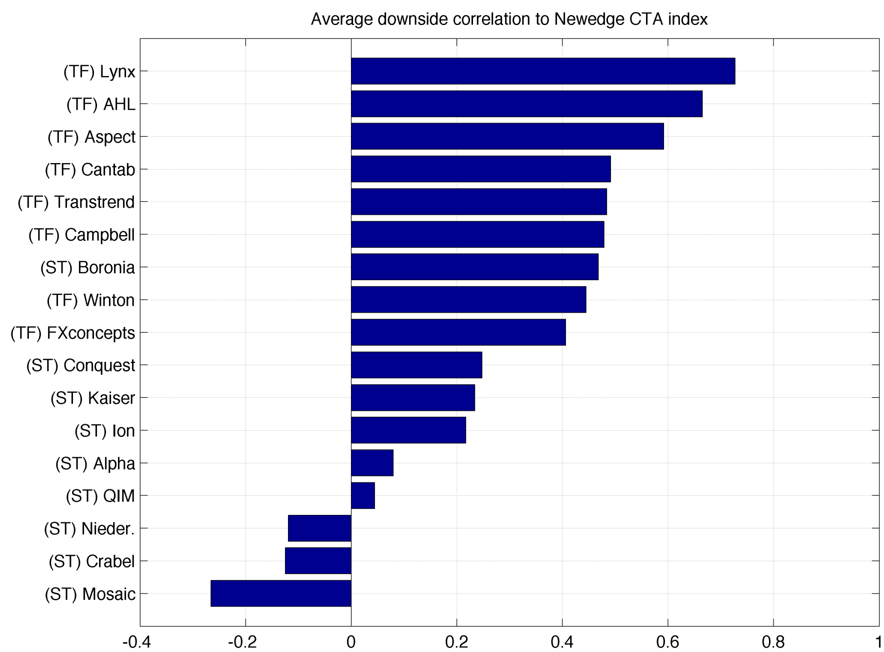

Figure 19: Downside correlation to the Newedge CTA Index (March 2007 - November 2012). The correlation is computed only for months when the Newedge CTA Index is negative.

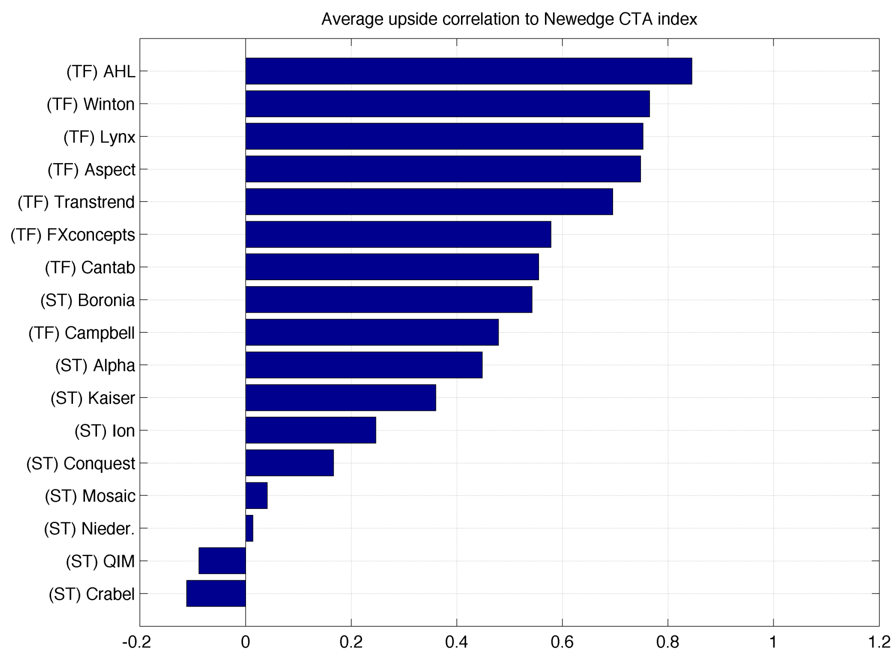

Figure 20: Upside correlation to the Newedge CTA Index (March 2007 - November 2012). The correlation is computed only for months when the Newedge CTA Index is positive.

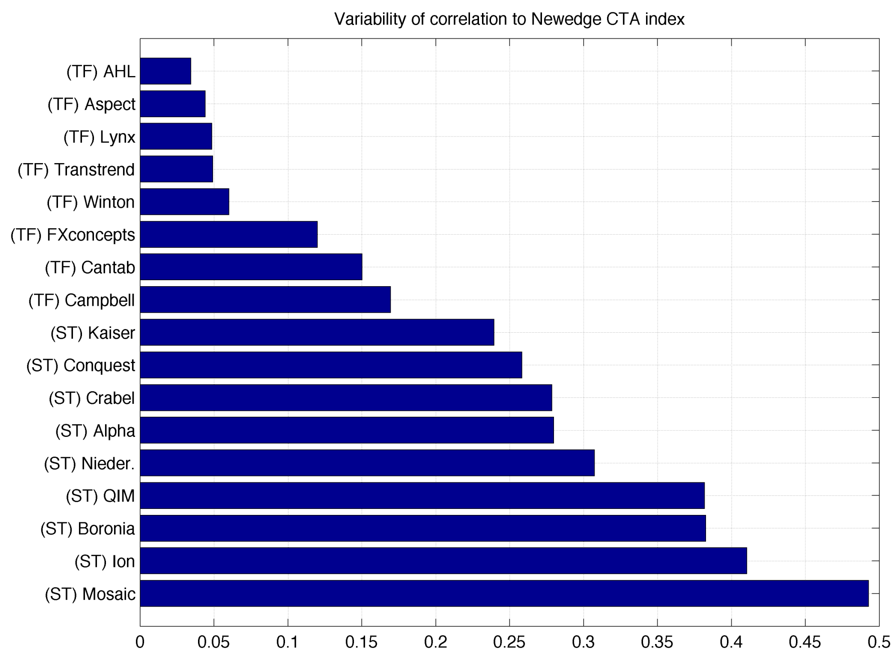

Figure 21: Variability of correlation to the Newedge CTA Index. The values are the standard deviations of the running 12-month correlations to the Newedge CTA Index.

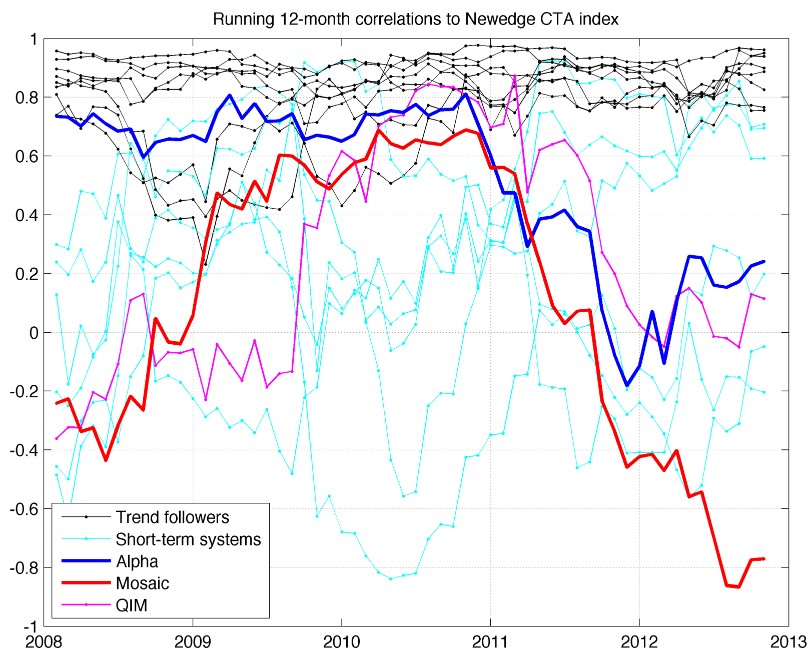

Figure 22: Time history of running 12-month correlations to the Newedge CTA Index. TF systems are shown in black, ST systems are shown in cyan, and other systems are shown in various colors.

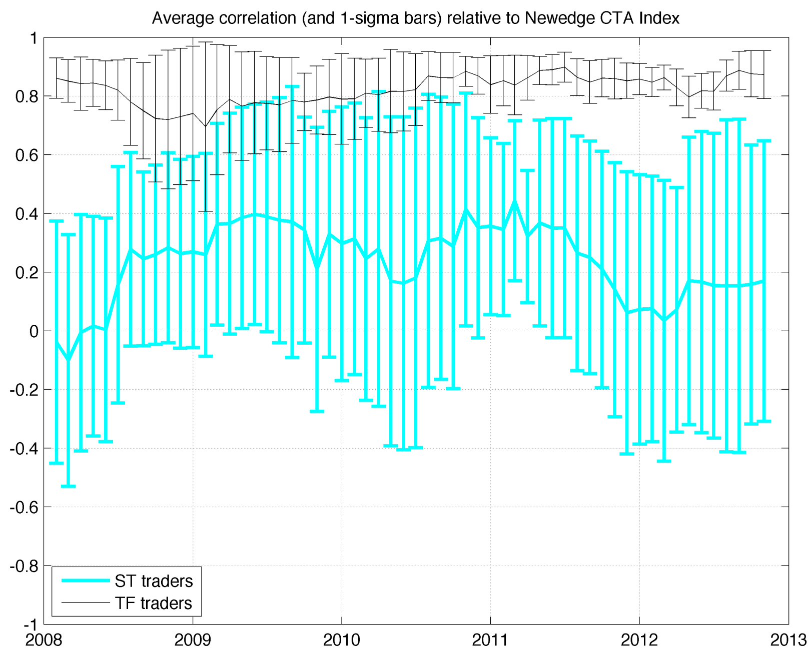

Figure 23: Average values of running 12-month correlation, separated out by TF programs (black) and ST programs (cyan). The error bars are +/- 1-σ values.

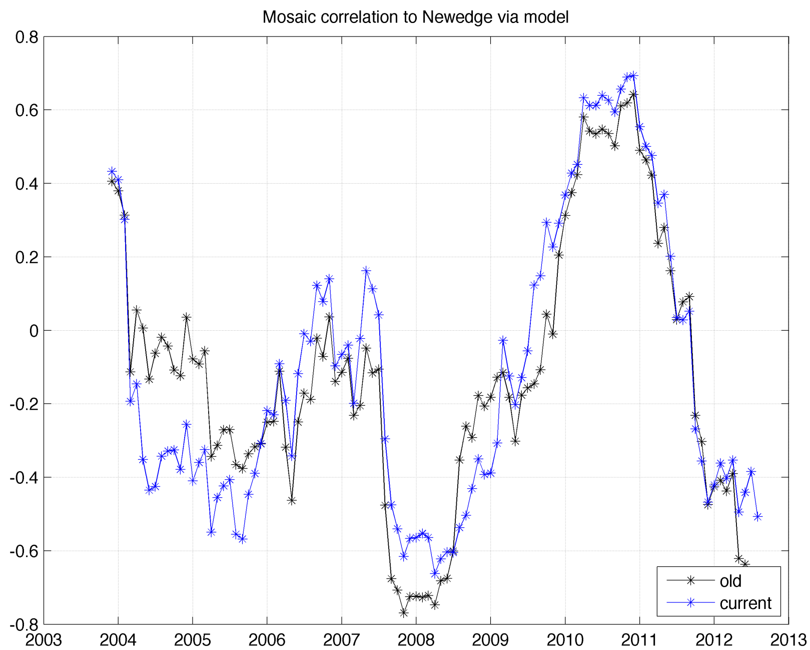

Figure 24: Modeled running 12-month correlation between Mosaic and the Newedge CTA Index. The system shows a propensity for the correlation to vacillate; this has become more pronounced since 2007. Recent updates to the model weights (black curve vs. blue curve) do not appreciably alter the results.

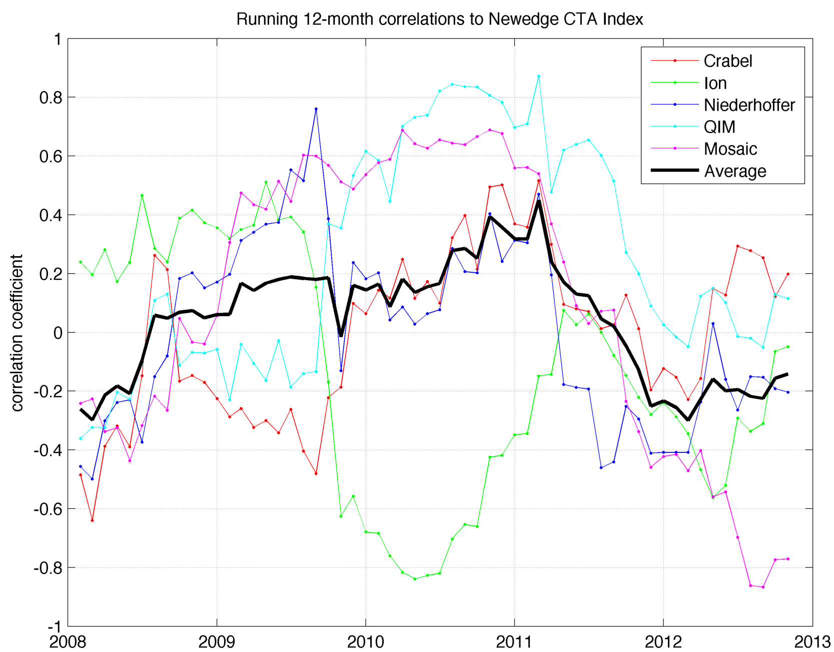

Figure 25: Running 12-month correlations between non-trend-biased short-term traders and the Newedge CTA Index. The black curve is the average.

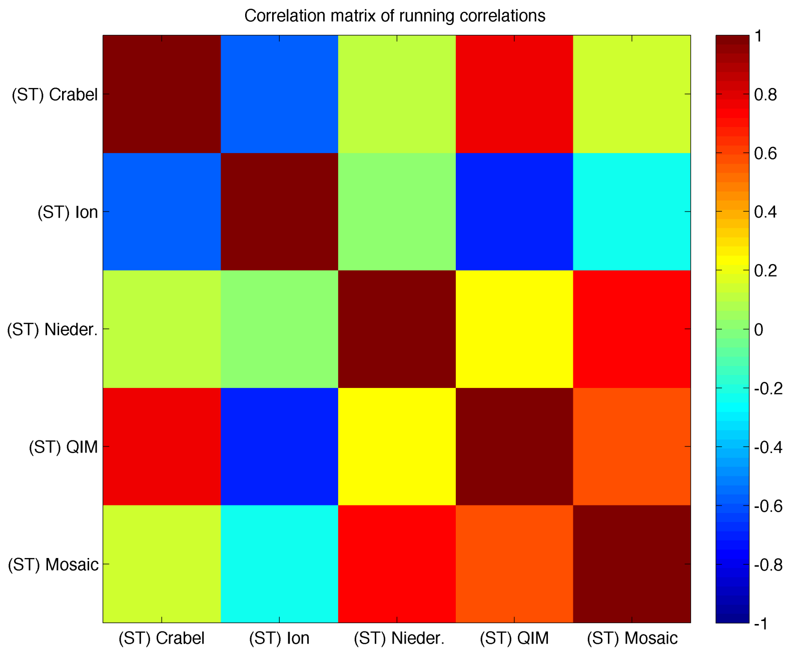

Figure 26: Correlation coefficients of curves shown in Figure 25. A negative correlation means that one system has a high correlation to the Newedge CTA Index when the other is low and vice versa. A positive correlation means that their correlation to trend following waxes and wanes synchronously.

For N completely uncorrelated systems where the pairwise correlation is always 0, we have the relationship

$$D = \sqrt{N}, \tag{3}$$

which matches the previous equation for $N \to \infty$ (and $C = 0$ as is already assumed). The critical item to note is that, in the above equation, we are contemplating N completely uncorrelated systems, which is the same as saying that we have N degrees of freedom. If we substitute the verbiage "DOF" for N and then combine the two equations, we can write

$$DOF = \frac{1}{C}, \tag{4}$$

which effectively now predicts the effective number of degrees of freedom given the average pairwise correlation of the systems being considered (and assuming N≈5 or more so that we can approach the asymptote). There are a couple of caveats in this case, the first being that frequency of representation matters. If most of the systems being considered are trend followers, the average correlation will rise to reflect the extra weighting given to trend dynamics. Second, time scale also matters in terms of minimizing the insidious effects of idiosyncratic noise. We find that the 1- to 3-month time-scale range is optimal for reducing idiosyncratic noise without introducing the potential performance-converging effects of longer-term, macro-economic drivers that we consider in Section 7.

The program correlations on monthly and 3-month time scales were shown in Figures 5 through 7. These reveal a solid expectation of how each program can be expected to behave over time (assuming, of course, no changes are made to the programs during the sample period). However, we are additionally interested in the "micro-structure" of the programs' similarities both to each other and also to aggregate indices (e.g. the Newedge CTA Index). To facilitate this goal, we need to analyze the data more deeply. Figure 17 shows the average correlation of each program to the Newedge CTA Index. Not surprisingly, the TF-system correlations exceed those of the ST programs. There is also a rough division within the short-term systems. Alpha, Kaiser, and Boronia all have roughly a 0.55 correlation to trend-following and thus bridge the gap between the long-term trend followers and the non-trend-biased short-term systems. The remainder of the short-term programs exhibit correlations between -0.05 and +0.32 to the Newedge CTA Index. Figure 18 shows the average correlation of each program to the others. The relative rankings are quite similar to the previous figure, indicating that trend-followers are correlated to each other but short-term systems are largely distinct on 1-month time scales. This is consistent with our DOF determination in Section 5. The 5 most unique systems in this study are Niederhoffer, Mosaic, QIM, Ion, and Crabel.

Figures 17 and 18 are somewhat standard fare for analytics, but now we move on to less-traditional metrics. Figures 19 and 20 break down the correlation to the Newedge CTA Index.

Specifically, we look at "downside" and "upside" correlation separately. Downside correlation includes only data where the Newedge CTA Index is negative, while upside correlation only correlates those months where the index is positive. This helps one determine if there is a nonlinearity to the correlation that might make a system more or less desirable than the average correlation would suggest. For the TF systems, there is no observed nonlinearity. However, for several of the short-term systems, the upside correlation is positive while the downside correlation is negative. The nonlinearity may be desirable in a portfolio where downside risk mitigation is of primary importance and upside correlation across systems is tolerable. Much of our own recent research, in fact, has focused on the development of intra-day models that can help increase this type of nonlinear, "V-shaped" correlation to trend-following returns.

We next turn to the **variability** of correlation. This is arguably as important as the assessment of average correlation values, yet it is far more difficult to interpret. Figure 21 shows each program's variability of correlation relative to the Newedge CTA Index. These values are obtained by computing a time series of 12-month running correlations and then computing the standard deviation of this curve for each system. TF systems, partially because they are in the index itself and partially because all the systems are so similar, show little variability. Short-term systems, however, show about 4 to 5 times as much variability, on average, with Mosaic exhibiting the largest deviations. Figure 22 shows the individual time histories of the 12-month correlations, while Figure 23 shows the time evolution of the TF and ST averages and variability. One curious observation is the range of Mosaic's correlation to trend following. Especially in 2012, Mosaic has continued to exhibit a very negative correlation to trend-following even as the other systems began to show increasing correlation around Q2 2012 to Q3 2012. To confirm both the historical range of Mosaic's expected correlation as well as the recent behavior, we examined simulation results extending back to 2003. Figure 24 shows the results for both the previous and current Mosaic systems (the only difference being that the latter has a reduced exposure to counter-trend models while keeping the average trend-following correlation nearly the same). The figure shows that both model versions behave similarly and produce a large variability in running 12-month correlations to the Newedge CTA Index (the range being approximately +/- 0.7). Moreover, the previous low in late 2007/early 2008 produced a similar negative correlation to what we observe now. So although Mosaic appears to have a wider-than-average range of correlation, this is not historically abnormal for the system. Since we know that the primary difference between Alpha and Mosaic is the presence of counter-trend models in the latter, we can attribute the correlation excursions to Mosaic's exposure to such models.

Even though we know the exact construction of our systems, we typically focus on targeting an **average** correlation to trend-following. Hence, the variability that we see across short-term systems in general is a bit surprising even to us. However, we can provide an educated guess as to the root cause of the correlation variability. First, many short-term systems, Mosaic included, incorporate multiple strategies, and each strategy often operates on a different time scale. Depending on the market environment, one type of strategy may be much more active than another. Figure 25 shows that in environments where trend-following is profitable (i.e. 2008, 2010), the correlation is typically rising. When trend-following is unprofitable or flat (2009, 2011, 2012), the figure shows flat to declining correlations to trend following (on average). This suggests that even non-trend-biased, short-term systems typically experience coincident signaling during trendy periods while exhibiting more diverse signaling during non-trendy periods. Even this simplifies reality a bit, as Figure 26 shows. This plot shows the correlation matrix of the running correlation curves in Figure 25. Essentially, this plot reveals which systems see their correlation to trend following rise and fall in tandem (or not). Mosaic, for example, exhibits correlation fluctuations that look much like Niederhoffer and QIM but not Ion or Crabel. QIM exhibits a high correlation to Crabel as well, likely reflecting the complex behavioral overlaps unveiled with the principal component analysis in Section 5.3.[3]

In aggregate, we observe that non-trend-biased short-term systems exhibit a highly-variable correlation to trend-following returns. The evolution of the variability is largely unpredictable but correlation generally appears to increase when trend-following returns are consistently positive. What appears less random is the "correlation of correlation", i.e. the tendency of different short-term systems' TF correlation to move in sync with or opposite each other. For example, Crabel and Mosaic show strongly opposing tendencies in 2009 and 2012; in 2009, counter-trend systems did very well and in 2012 they did quite poorly. This suggests that Crabel's systems oppose such dynamics, while Mosaic clearly exploits them. More importantly, it suggests ways in which to identify complementary short-term systems that have the potential to mitigate each other's risk across a wide range of market environments. This has important ramifications for portfolio construction.

7 Stress testing and stressor-response-based portfolio construction

We have introduced several methods by which to move beyond "efficient-frontier-like" optimization. Some of these, like PCA, are already utilized in certain instances, and others, such as correlation considerations, are often employed in their simplest form to make global measurements independent of time scale or downside/upside delineations. We believe that our extensions to these metrics have potential merit in creating portfolios with better risk-adjusted performance and higher temporal consistency. However, even the use of these techniques may well have been insufficient to offset the primary challenge that futures-based, multi-strategy portfolios have experienced in the past four years: as Figure 2 shows all too well, there is a tendency for long-time-scale returns of our candidate short-term and long-term systems to overlap. In 2007 and 2008, both types of systems did well, and

[3] As a brief aside, we highlight three systems in Figure 22: Mosaic, Alpha, and QIM. Looking at the PCA results from Section 5, we found that QIM and Mosaic have historically been quite similar (especially with respect to PC #3). When we first looked at the running correlations, we were surprised at the recent divergence between QIM and Mosaic. We see a phenomenon that we subsequently confirmed by comparing Mosaic's returns to QIM's and then Alpha's to QIM's. There appears to have been a shift in QIM's strategy from one similar to Mosaic to one more similar to Alpha, i.e. counter-trend models have been reduced in importance relative to trend-reversion or trend-following models. This seems to have happened sometime in early- to mid-2011. We can also see the overabundance of counter-trend models in QIM relative to Mosaic during 2009 (since our level of counter-trend exposure was increased in December of 2009, this makes complete sense). Since 2011, QIM's correlation to the Newedge CTA Index has much more closely followed Alpha. We do not know if this strategy migration was a result of capacity-driven, trading-cost considerations or re-optimization of the system mix (or something else), but it appears to have been fortuitous timing given Alpha's and QIM's outperformance of Mosaic since that time.

	Low correlation environment		Average correlation environment		High correlation environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
ID CCT 16x	-0.60	-47.4, -18.1	-0.09	-4.9, 2.2	0.86	62.1, 11.7
ID CT 16x	0.67	24.4, -31.5	0.03	10.5, -2.7	-0.75	-56.0, 39.8
Newedge CTA Index	0.24	30.0, -0.7	0.14	20.4, 0.9	-0.66	-91.3, -1.9
Internal TF	0.67	50.5, -9.0	0.19	19.9, -0.6	-1.20	-110.2, 10.8
Alpha EOD	0.73	23.8, -14.1	0.33	21.3, -0.6	-1.72	-87.6, -15.9
Mosaic EOD	0.88	21.8, -18.6	-0.05	-1.2, -1.7	-0.73	-18.2, -23.7

Table 2: Performance versus correlation values.

	Low correlation environment		Average correlation environment		High correlation environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
TF 2x	1.45	95.3, -15.7	-0.04	2.1, -2.7	-1.32	-101.5, 23.8
TF 4x	0.39	28.9, -13.1	0.33	37.4, -2.5	-1.37	-141.2, 20.5
TF 8x	0.68	67.0, -8.7	0.17	22.6, -1.1	-1.18	-134.8, 12.1
TF 16x	0.44	60.8, -3.9	0.09	13.3, -1.8	-0.70	-100.7, 9.4
CT 2x	-0.37	-102.1, -23.7	-0.27	-83.1, -2.8	1.19	351.5, 32.2
CT 4x	0.23	-5.3, -17.0	-0.49	-67.6, -5.9	1.25	208.2, 34.9
CT 8x	1.45	57.7, -27.9	-0.28	-17.7, -8.2	-0.61	-4.5, 52.3
CT 16x	1.68	48.5, -38.7	-0.23	-1.5, 0.4	-1.00	-44.1, 37.4
CCT 2x	2.19	64.6, -20.1	0.06	9.5, -0.7	-2.38	-93.1, 22.1
CCT 4x	1.94	44.8, -23.2	0.13	12.7, -0.5	-2.33	-83.0, 24.7
CCT 8x	0.06	-19.7, -23.3	0.26	10.6, 0.6	-0.83	-12.0, 21.6
CCT 16x	-1.29	-77.2, -17.9	0.24	8.9, 0.2	0.57	50.4, 17.2

Table 3: Performance versus correlation values (ID models).

since then, both have been net flat to negative. Unlike the structurally-mandated de-correlation of stocks and futures (due to the ability of futures-based strategies to profitably "short" the economy during bear markets), there is no obvious structural dichotomy between short-term and long-term futures-based systems. Nonetheless, it is worth exploring the possibility that such dichotomies exist between certain subsets of possible trading styles. If one could find and understand such divergent behavior, this could be profitably exploited by constructing a portfolio that is more likely to perform in a broad cross-section of macro-economic environments. Even though one can logically and persuasively argue that futures-based strategies have done what they are designed to do in the post-financial-crisis environment (2009-2012) and that a stock/bond/futures mix still appears to have the best "all-weather" performance, there is an additional challenge that must be overcome. That is the finite and fairly short (e.g. 2-4 years) patience of investors to endure flat to slightly-negative net returns in a given strategy, however illogical it may be to implement such punitively-tight time constraints.

As an alternative to correlation studies, which doesn't address the *cause* of positive or negative periods of returns, we decided to focus on how certain macro-economic environments may influence the performance of various strategies. To that end, we have done a considerable amount of research examining whether we can identify certain types or styles of strategies that do especially well (or poorly) in response to specific, well-defined stressors. Although this falls short of truly identifying particular macro-economic environments, it does at least focus on behaviors that might be readily associated with particular environments. We looked at five specific, quantitative metrics and then also looked at the rate of change of these metrics in order to arrive at 10 different "stress conditions" or stressors:

1. Average cross-market correlations.
2. Rate of change of cross-market correlations (i.e. is the value increasing, decreasing, or constant?).
3. VIX value (the well-known Volatility Index).
4. Rate of change of VIX.
5. Price divergence on 4-day to 8-day time scales. This is essentially a measure of how oscillatory price movements are on these time scales.
6. Rate of change of the 4 to 8 day price divergence.
7. Price divergence on 16-day to 28-day time scales.
8. Rate of change of the 16 to 28 day price divergence.
9. Price divergence on 40-day to 50-day time scales.
10. Rate of change of the 40 to 50 day price divergence.

	Low Vix environment		Average Vix environment		High Vix environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
ID CCT 16x	0.37	25.2, 3.8	-0.13	-11.7, -4.4	0.02	10.0, 9.4
ID CT 16x	-0.05	-8.3, -3.8	0.09	8.2, -1.3	-0.22	-16.3, 7.6
Newedge CTA Index	1.13	116.5, -18.3	-0.59	-80.3, 1.3	0.65	124.5, 14.5
Internal TF	1.28	76.2, -18.2	-0.76	-75.0, -4.0	1.02	148.9, 30.1
Alpha EOD	1.10	53.7, -4.2	-0.33	-18.2, -1.2	-0.13	1.0, 7.9
Mosaic EOD	0.67	19.1, -11.7	-0.08	-3.4, -1.4	-0.43	-8.8, 16.0

Table 4: Performance versus Vix values.

Stressors 1-4 and 5, 7, and 9 are fairly easy to interpret. With respect to the latter 3, if prices are oscillating on particular time scales, then there is by definition a lack of trendiness and a large amount of mean reversion. A high-level of price divergence is thus synonymous with oscillatory behavior on that time scale. Items 6, 8, and 10 are a bit less intuitive, but they tend to correlate highly with items 5, 7, and 9 respectively. Thus they can be considered redundant measures for our purposes.

Table 2 shows sensitivity results for the cross-market correlation stressor (item 1). We divide the range of correlations into five quintiles: the lowest quintile is designated the "low-correlation environment" and the highest quintile is designed the "high-correlation environment". The middle 3 quintiles (60% of the data) are considered the "average correlation environment". We show results for the Newedge CTA Index, an internal long-term trend-following model, our Alpha program (EOD models only), our Mosaic program (EOD models only), a high-frequency trend-reversion ("CCT") system, and a high-frequency mean-reversion ("CT") system.[4] Focusing on the "delta Sharpe" value, we note that almost all of the systems perform relatively better in a low-correlation environment that a high-correlation environment. The lone exception is the CCT 16x model.

Based on this interesting result, we then repeated the study with intra-day models only. The results are shown in Table 3. Note that the values for CT 16x and CCT 16x are slightly different than in the previous table because the time period sampled was not identical between the two tables. However, what is clear is that there is a strong time-scale dependence on the sensitivities. As the trading frequency is increased, intra-day TF models improve slightly in a high-correlation environment while intra-day CT models degrade. CCT models, on the other hand, improve tremendously in a high-correlation environment as the trading frequency is increased, thus highlighting the dramatic dependence that time scales can have on performance even when the model construction is unchanged. This finding also supports the analysis performed in Section 5.1.

Repeating this analysis for the other stressors eventually reveals a sizable challenge for the portfolio designer. It should be first noted that divergence generally causes a wide spread of responses because it is time-scale oriented. Thus, what is bad for Alpha and Mosaic is often good

[4]We require daily data for these analyses, so this precludes any comparisons using external programs.

	Low correlation ROC environment		Average correlation ROC environment		High correlation ROC environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
ID CCT 16x	-1.56	-96.1, -11.8	-0.89	-60.1, -1.6	4.22	276.4, 16.6
ID CT 16x	1.48	154.4, -18.4	1.11	144.8, -10.8	-4.81	-588.6, 50.9
Newedge CTA Index	0.14	20.3, -8.1	0.78	107.8, -2.8	-2.47	-343.6, 16.5
Internal TF	-0.09	-3.8, -7.2	0.90	80.8, -4.6	-2.59	-238.7, 21.1
Alpha EOD	0.65	28.8, -13.0	1.38	81.7, -3.0	-4.78	-274.0, 21.9
Mosaic EOD	0.10	4.7, -8.4	1.32	71.1, -3.4	-4.06	-218.1, 18.6

Table 5: Performance versus correlation rate-of-change (ROC) values.

	Low Vix ROC environment		Average Vix ROC environment		High Vix ROC environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
ID CCT 16x	-1.96	-111.3, 0.1	0.35	14.9, -3.2	0.92	66.5, 9.5
ID CT 16x	-0.72	-55.6, 17.1	0.99	96.8, -9.6	-2.26	-234.7, 11.8
Newedge CTA Index	-2.42	-344.0, 7.3	0.84	105.9, -10.7	-0.10	26.4, 24.9
Internal TF	-3.91	-374.2, 11.0	0.99	69.5, -14.8	0.94	165.7, 33.5
Alpha EOD	-1.94	-100.1, 6.9	0.87	39.8, -7.5	-0.65	-19.4, 15.6
Mosaic EOD	-0.16	11.8, 16.6	0.43	10.5, -11.8	-1.11	-43.3, 18.9

Table 6: Performance versus Vix rate-of-change (ROC) values.

	Low divergence environment		Average divergence environment		High divergence environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
ID CCT 16x	-0.07	5.2, 7.1	-0.23	-16.4, -4.5	0.76	44.0, 6.3
ID CT 16x	0.12	16.5, 6.2	-0.23	-24.5, -3.2	0.57	56.9, 3.5
Newedge CTA Index	0.83	118.7, -5.9	0.15	26.7, 0.5	-1.29	-198.9, 4.3
Internal TF	0.43	48.3, -2.9	0.21	24.2, 1.0	-1.04	-120.9, -0.1
Alpha EOD	1.20	90.8, -0.45	-0.48	-37.1, -0.1	0.24	20.4, 0.7
Mosaic EOD	1.05	48.6, -4.7	-0.29	-22.1, -7.8	-0.19	17.6, 28.0

Table 7: Performance versus 40-50 day divergence values.

	Low Vix ROC environment		Average Vix ROC environment		High Vix ROC environment	
Model	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma	Delta Sharpe	Delta mean, sigma
TF 2x	-3.43	-315.5, 14.6	0.70	46.5, -12.9	1.32	176.1, 24.1
TF 4x	-4.27	-501.5, 16.8	0.75	55.4, -17.6	2.03	335.4, 35.9
TF 8x	-4.75	-615.3, 11.7	0.54	33.8, -15.2	3.13	513.8, 33.8
TF 16x	-3.66	-667.2, 10.8	-0.17	-60.7, -13.2	4.17	849.2, 28.9
CT 2x	2.22	585.1, 20.2	0.03	-1.0, -17.8	-2.30	-582.1, 33.2
CT 4x	2.46	352.4, 21.6	0.19	7.3, -16.3	-3.04	-374.1, 27.3
CT 8x	0.66	96.7, 23.5	0.71	38.1, -13.6	-2.78	-211.1, 17.2
CT 16x	-0.77	-32.4, 17.8	1.18	80.0, -11.5	-2.78	-207.7, 16.9
CCT 2x	-1.58	-60.6, 10.8	0.74	22.9, -9.4	-0.62	-8.2, 17.5
CCT 4x	-3.11	-125.2, 12.6	1.11	37.9, -9.5	-0.22	11.6, 15.8
CCT 8x	-3.25	-116.7, 3.6	1.00	28.1, -7.5	0.25	32.4, 18.9
CCT 16x	-2.12	-118.9, 0.4	-0.05	-8.5, -3.5	2.27	144.3, 10.0

Table 8: Performance versus Vix ROC values (ID models).

for trend-following and vice versa. These stressors tend to represent different states of normal market dynamics, and a proper combination of longer-term TF models and shorter-term non-TF models can reduce the sensitivity to these "normal" market variations. Table 4, for example, shows the response to the VIX stressor[5]. In this case, we see a desirable result, which is that intra-day models are fairly neutral, trend-following models outperform in high-VIX situations, and Alpha and Mosaic underperform when the VIX is high. Thus, by mixing models carefully, we can reduce the sensitivity to extreme moves in the VIX (both low and high). Looking back at Table 2, however, we see that **none** of the end-of-day models we sampled show a positive response to a high-correlation environment. Table 5 shows the corresponding results for the correlation rate-of-change (ROC) stressor, and the results are even worse. In environments where correlation is climbing rapidly (typically during a market panic), **all** end-of-day models strongly underperform on a relative basis, and Alpha and Mosaic are worse than trend-following systems. Next, Table 6 shows the results for the VIX ROC stressor. In this case, TF models are fairly neutral in a high VIX ROC environment and help counteract the negative performance sensitivity of Alpha and Mosaic. Finally, in order to provide some intuitive feel for the divergence stressor, Table 7 shows results for the 40-50 day divergence metric. If prices are divergent on this time scale, it implies a large amount of oscillatory behavior (i.e. reversals). Since this time scale coincides with that of trend followers, we would expect large divergence values to be bad for trend following, and this is exactly what the results show. Intra-day models, on the other hand, tend to relatively outperform in this environment, and our end-of-day Alpha and Mosaic models tend to be neutral with respect to this environmental condition. What we generally see is that, in cases where no known EOD

[5]After this analysis was finished, we came across a similar study in reference [6]. Exhibit 9.10 in this book shows qualitatively similar results using aggregate CTA returns (i.e. primarily trend followers).

models operate profitably in a given environment, models of higher frequency (i.e. high-frequency ID models) tend to counteract these losses. This is an expected result if our time-scale cascade argument from Section 5.1 is fundamentally correct. In this case, by moving to vastly shorter time scales, we can take advantage of the same relative dynamic that is unprofitable on longer scales.

Aside from diagnostics, these sensitivities have potential use in portfolio optimization. We have tested the idea of constructing portfolios using weights such that the total sensitivity to different market environments is minimized. In other words, strategy weights are chosen such that no single stressor (at least those that we consider) causes an expectation of outsized positive or negative performance. The portfolio is thus selected to be as "stressor neutral" as possible. Using only this constraint, the results show a propensity to find systems with very good consistency over time and nearly-neutral skew and kurtosis (both downside and upside). Thus, without appealing to performance or correlation directly, simply optimizing to "regularize" stressor responses appears to inherently produce statistically-robust portfolios.

The challenge, of course, is to generate a list of stressors that covers all realistic situations. We don't claim that our list achieves this goal. Our choice to focus on correlation spikes and equity-market volatility is obvious, as these have been present in all the market panics of the last several years. Additionally, the divergence measures tend to cover a wide gamut of "normal" market gyrations, so these help round out our efforts to represent the "state of the world" using a finite, small number of metrics. However, there are likely other measures of stress that we haven't identified; the challenge to the portfolio designer is conceiving of and quantifying them, and then ensuring that their presence is sufficiently repetitive that robust statistics can be generated around them. Some software packages allow one to measure responses against actual historical events such as the Lehman collapse, the Gulf Wars, and the Asian crisis of 1997 (among others). This is not an avenue we have yet explored, but benchmarking our results against specific, well-defined events would be a useful exercise to gauge the consistency of system responses versus our somewhat-general stressor conditions. Thus, we view these approaches as highly complementary and certainly worth additional exploration. Another possibility is to use synthetic price data to gauge stress responses.[6] The difficulty of this is that it requires a deep level of coordinated effort between the portfolio designer and the candidate CTAs. However, it also potentially allows the most powerful method of gauging and then maximizing overall portfolio robustness. If the portfolio designer is able to create synthetic data that includes correlation spikes, volatility spikes, trend events, reversal events, and other anomalies of varying severity, and if the CTA is able to treat this data as "real" and hence generate simulated per-market profit/loss, the designer would truly be able to arrive at detailed estimates of true portfolio risk. It would also much more clearly show which programs are structurally-decorrelated and which others may tend to implement conditionally-correlated strategies whose behavior coincides under certain conditions.

While we present this as a fairly-speculative and perhaps radical idea, it nonetheless has great promise as an alternative to mean-variance optimization. After all, mean-variance optimization necessarily looks at average risk/reward tradeoffs; moreover, it is "naive" in that it doesn't delve

[6] Synthetic price data can be generated fairly easily using T-distributed data with varying degrees of freedom and a simple mechanism to cluster both volatility and cross-market correlation in time. Moreover, the level of fat tails in the price data can be very finely controlled in order to generate results ranging from the fairly-likely to the seemingly-impossible.

in the "whys" of the returns. Conversely, the ideas presented here focus on measuring and understanding specific responses to specific environments. On the one hand, it can optimize for consistency (essentially, the *variability* of risk/reward), and on the other, it can also focus narrowly by conditioning to certain market environments or events (the *tails* of risk/reward). And instead of modifying the mean-variance approach with various constraints or penalty functions to secondarily incorporate these concepts, it instead uses them as the **primary** optimization targets, with the expectation that the average risk/reward will automatically be pseudo-optimized as well. At the very least, our strong belief is that the approaches should be used in tandem to allow the designer more information with which to make critical portfolio-allocation decisions.

8 Conclusions

This paper has explored some alternative methods by which to analyze and compare the properties of trading returns from different programs. We have focused specifically on systematic CTA strategies of two main types: trend-followers and short-term traders. We first showed that short-term kurtosis tends to be aligned with trading style and is not necessarily due to the skill of the program designer. Short-term traders with momentum strategies, not surprisingly, tend to show the greatest propensity for positive fat tails in one-month returns. However, on three-month time scales, there is a good deal of mean reversion. Mosaic, while explicitly avoiding momentum-based strategies, is designed to approach "normality" of returns as quickly as possible and hence shows close to 0 excess kurtosis on three-month time scales. In short, risk management in large part depends on the strategies being implemented and not primarily the skill (or lack thereof) on the part of the program manager. Thus, as we argued in a previous paper [2], optimization routines that heavily reward positive skew may inadvertently cause "inbreeding" by causing selection only of similar strategies (that may nonetheless look somewhat diverse due to idiosyncratic noise).

We next considered the true number of independent return streams from which the portfolio manager can choose. Applying a number of different techniques, we arrived at the conclusion that there are likely only 4 to 6 true DOF available. Even this sobering result may paint a too-optimistic picture of diversification potential, as non-trend-biased strategies have shown a longer-term correlation to trend-based programs during the past four years. On the other hand, even though idiosyncratic noise does not represent true DOF, it can be exploited in order to smooth returns and more efficiently extract the underlying alpha.

Following this, we looked at correlation in relatively-unorthodox manners: conditionally (e.g. upside vs. downside relative to the Newedge CTA Index), over time, and also in terms of its variability and how this depends critically on the program's trading style. We find that short-term systems have more variability than TF systems, but careful analysis of the phase relationships in short-term-systems' correlations to trend-following may help the designer mate complementary programs to optimally augment a trend-following core.

Finally, we touched on an alternative portfolio-optimization technique thats avoid performance as the primary objective and instead focuses on macro-environmental and market-stress responses. Our own experiments with this approach show that performance arises as a natural outcome of such optimizations even though it is explicitly avoided as a goal. Since the primary goals of consistency and robustness are also necessarily satisfied in these exercises, this appears to be an approach with great potential for improving upon current methods.

We want to share two additional thoughts before closing: first, quality still matters. The theme of this paper has been about diversification and identifying what is truly "different". However, even to the extent that a number of programs are the "same", there are still factors that might make one preferable to another. Simply because two programs exploit the same underlying dynamic doesn't mean that a particular manager can't exploit it more efficiently than the others. Thus, the results presented herein shouldn't be construed to imply that return streams are a commodity, only that the available diversification is finite and limited.

Second, we noted that most available degrees of freedom necessarily exist at short time scales (10 days or less). While short-term programs don't guarantee true (as opposed to idiosyncratic) diversification, it is still the most promising place to look for the needed diversification. However, we also showed that short-term programs often have complex overlaps of trading styles. As an example, if we assume that short-term systems have 3 DOF labeled A, B, C, then Program 1 might have 50% A, 30% B, and 20% C. Program 2 might have 0% A, 30% B, and 70% C. Program 3 might have 25% A, 50% B, and 25% C, and so on. Thus, correlations between programs 1, 2, 3, etc. will be quite hard to discern as they will look different in different market environments. Principal component analysis and running correlation analysis can expose these complex overlaps to some extent, but there is no substitute for working more closely with the program designers. Trend-following programs can, to a large extent, be offered as a "turnkey" solution, but we argue that this is not the case with short-term systems. It is imperative the the program manager and the portfolio designer understand each other's goals in a very intimate manner, and tools such as synthetic data-set analysis would help facilitate the deep understanding required to effectively conjoin a truly-diverse strategy set.

The tools discussed in this paper were largely developed as pragmatic responses to challenges we faced in our own program-building efforts. We have found that implementing non-trend-biased systems is difficult, and managing the risk associated with such systems is more difficult still. Finally, there is the question of preferred market environments. Many non-trend-biased systems outperform in relatively-benign market environments such as those pre-2008. The shock-prone markets witnessed in the past few years, however, have proven extremely difficult for nearly-all short-term CTA strategies of which we are aware. We have thus utilized the tools in this paper to analyze possible solutions to these challenges. We believe that they are applicable to a wide range of so-called "portfolios", (from intra-program to inter-program to inter-asset), and our hope is that these same tools can be refined and extended over time, thus providing a valuable payback to those willing to explore beyond the efficient frontier.

Appendix A - Testing artificial data with known degrees of freedom

In order to test the DOF estimation concepts used in Section 5, we created an artificial data set. The data set started with 4 true degrees of freedom that were generated from random, Gaussian data; we populated them with 69 samples each in order to match the lengths of the real, monthly

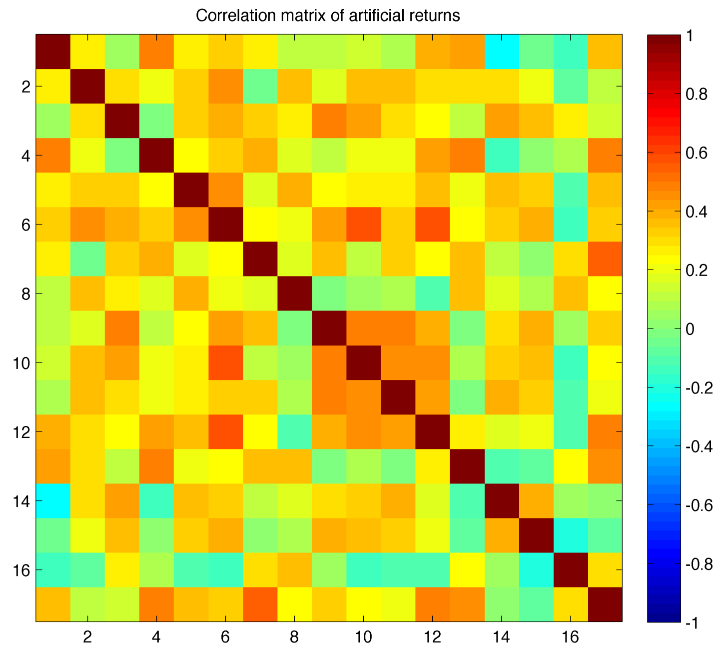

Figure 27: Correlation matrix for artificial data. The average pairwise correlation value is 0.23.

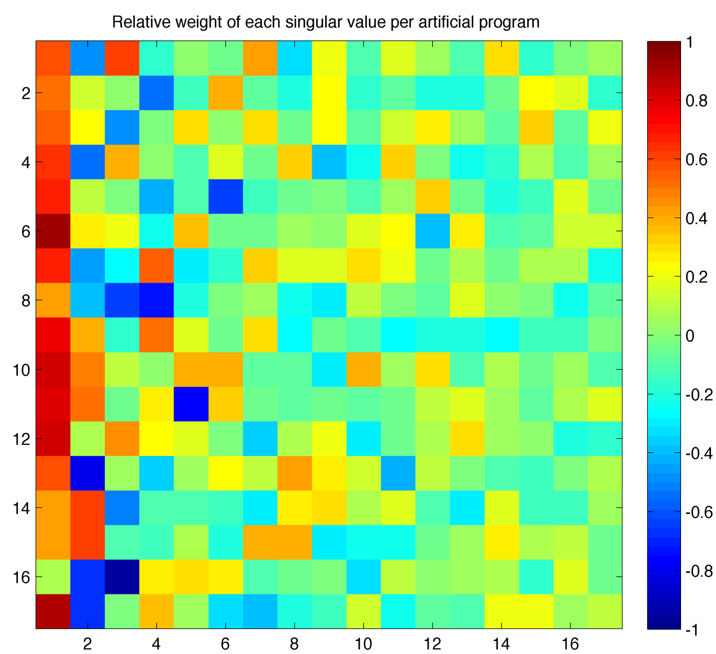

Figure 28: Amplitude of each principal component for each artificial program.

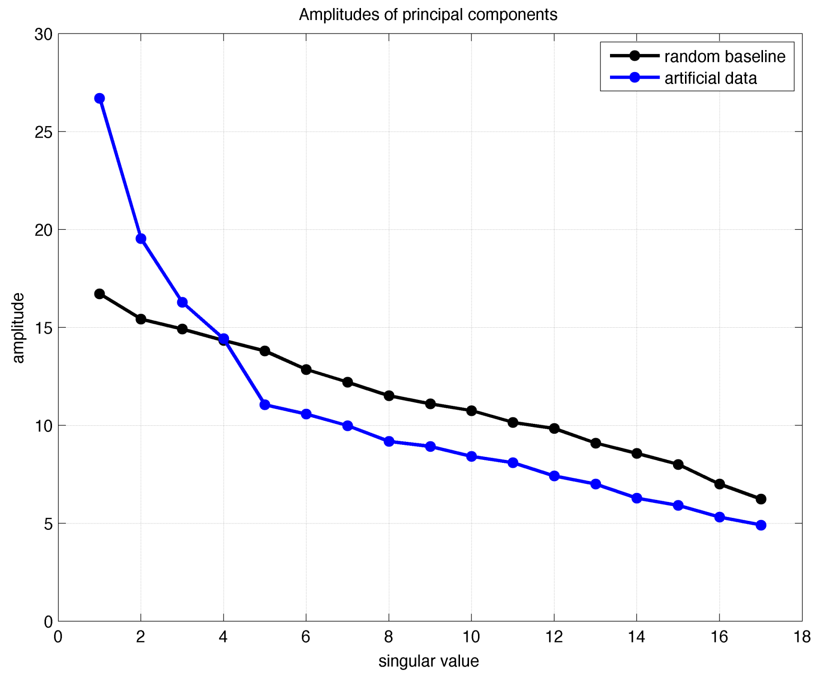

Figure 29: Principal component amplitudes for artificial system set in blue and random baseline in black.

data used in this paper. From these four independent strategies, we then generated a 17x4 matrix of random weights. For each of 17 artificial programs, we took four weights (i.e. one row of the weight matrix), multiplied each one by its corresponding DOF, and added them together to get a final set of artificial returns. In this manner, each of the artificial return sets consists of a randomly-weighted mix of four "strategies". In matrix form, this can be written as

$$\mathcal{R} \;=\; \mathcal{F}W^T, \tag{5}$$

where $\mathcal{R}$ is the 69x17 array of artificial returns, $\mathcal{F}$ is the 69x4 array holding the independent artificial strategies, and W is the 17x4 weight matrix. Finally, we add random noise to each return such that the standard deviation of the noise is about equal to that of the original returns. This mimics the idiosyncratic noise that we have discussed previously.

The correlation matrix of the final strategies is shown in Figure 27. The average pairwise correlation is 0.23, which is quite low and would be welcomed by most portfolio designers. However, since we constructed this set of strategies from four known DOF, we immediately recognize the expectation given by Equation 4, which is that 4 DOF should in fact yield a pairwise correlation = 0.25. Given the added idiosyncratic noise, there is some scatter around this value as the code for subsequent re-generation of the $\mathcal{R}$ array generates new noise values for each realization, but in general this relationship holds (as it should).

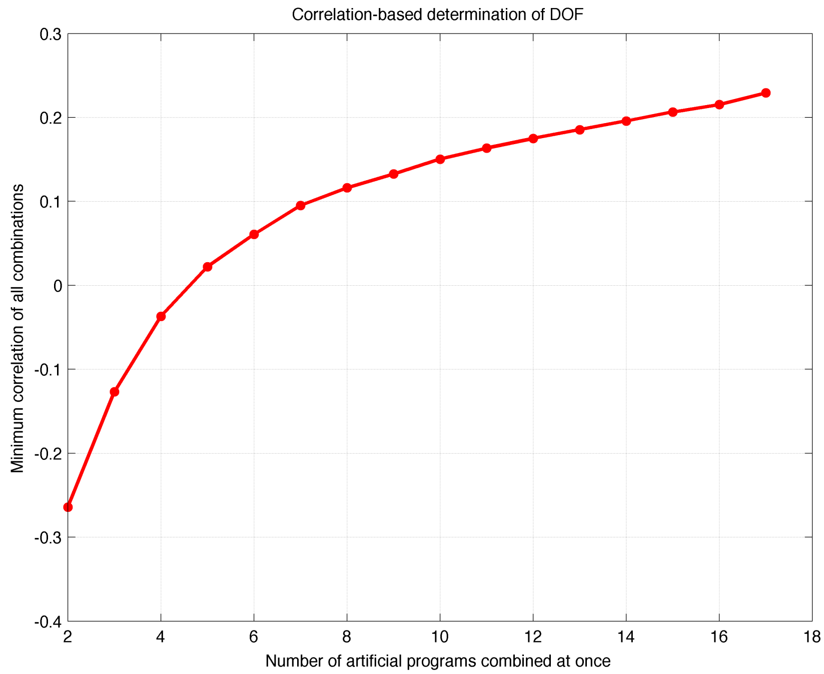

Figure 30: Minimum correlation obtained by taking N artificial programs at once, where N varies from 2 to 17.

We first perform the principal component analysis on the data set; the results are shown in Figures 28 and 29. These look remarkably similar to the decomposition of the real data. The first four principal component amplitudes are clearly above the noise floor, and it is a judgment call as to whether we would have determined the data set to have four or five DOF based on these results. We next use the correlation-based approach that was introduced in Section 5.2. The results of this estimation method are shown in Figure 30. The correlation first becomes positive when 5 programs are considered at once, thus indicating that there are indeed four DOF in the data set. To be fair, this result appears to be atypically accurate given the variability we saw when repeatedly re-running this analysis with random data. The curve is often moved up or down so that the zero crossover point is somewhat variable. We believe this is due to random fluctuations in what is quite a short data set. However, what does seem consistent is the point where this curve changes slope. If one takes the initial steeper slope and extrapolates to the right side of the graph, and then if one takes the shallower slope and extrapolates to the left side of the graph, these lines almost always intersect somewhere between 4 and 5 DOF, just as with the principal component analysis. Returning to Figure 8, this is an easy task for this realization given the linearity of the two line segments, and the result again indicates 4 DOF in the real data set.

To summarize, based on this simple proof-of-concept test, it does appear that the methods used in Section 5 have validity for estimating the number of independent return sources available to the portfolio designer.

References

[1] Mundt, Michael D., "The short-term dilemma", *Futures*, September 2011, 50-52.

[2] Mundt, Michael D., "The DNA of a diversified portfolio", *Futures*, December 2011, 50-52.

[3] Markowitz, Harry, "Portfolio Selection", *Journal of Finance*, March 1952, 77-91.

[4] Brigham, Eugene, Louis Gapenski, and Phillip Daves, *Intermediate Financial Management (Sixth Edition)*, Dryden Press, Fort Worth, Texas, 1999.

[5] Abrams, Ryan, Ranjan Bhaduri, and Elizabeth Flores, *Lintner revisited: A quantitative analysis of managed futures*, CME Group paper, www.cmegroup.com, 2012.

[6] Burghardt, Galen and Brian Walls, *Managed Futures for Institutional Investors*, John Wiley and Sons, New Jersey, 2011.

[7] Kaminski, Kathryn, "Diversify with crisis alpha", *Futures*, February 2011, 50-51.

[8] Rollinger, Thomas, "Revisiting Kat's managed futures and hedge funds: a match made in heaven", working paper, September 2012.

[9] Mundt, Michael D., "Understanding risk", *RCM internal paper*, 2011.

[10] Stone, James, *Independent Component Analysis: A Tutorial Introduction*, MIT Press, Cambridge, Massachusetts, 2004.