In Pursuit of Trend-Following Beta: The Promise and Pitfalls of Replication[‡]

STEVEN BRAUN, COREY HOFFSTEIN, JULIUSZ JABLECKI[1]
October 2024

Steven Braun is the Chief Derivatives Risk Officer, Portfolio Manager, and Senior Quantitative Analyst at Newfound Research LLC. 380 Washington Street 2nd Floor, Wellesley, MA 02481. E-mail: sbraun@thinknewfound.com.

Corey Hoffstein is the CEO and CIO at Newfound Research LLC. 380 Washington Street 2nd Floor, Wellesley, MA 02481. E-mail: corey@thinknewfound.com.

Juliusz Jabłecki is the Head of Risk and SAA at National Bank of Poland and Professor of Finance at University of Warsaw, E-mail: juliusz.jablecki@nbp.pl.

‡ Steven Braun is Chief Derivatives Risk Officer, Portfolio Manager, and Senior Quantitative Analyst at Newfound Research LLC. Corey Hoffstein is the Chief Executive Officer and Chief Investment Officer of Newfound Research LLC.

1 The authors would like to thank, in alphabetical order, Adam Butler, Conrad Ciccotello, and Chris Lundberg who offered their opinions and insights.

Title: In Pursuit of Trend-Following Beta: The Promise and Pitfalls of Replication

Date: October 2024

Abstract:

Against the background of a large body of research documenting the benefits of allocating to trend-following strategies, this paper identifies two related challenges facing potential allocators: mis-specifying and poorly executing a trend-following program oneself or selecting a manager who does the same. Both risks can overwhelm the diversification benefits of the strategy and so we examine whether they can be mitigated through replication of a broad index of trend-following funds. Using a series of numerical tests, we confirm that replication is indeed possible, both in an idealized case of a virtual fund of funds (via an ensemble of generic trend-following strategies) and in the more realistic scenario of an index composed of live funds (the BarclayHedge BTOP50 Index). However, we also show that simple regression-based replication introduces its own tracking risk driven by the trade-off between fee savings and skill leakage. We analyze the trade-off numerically and show that the effectiveness of replication can be improved through regularization techniques rendering it attractive as an active strategy that is able to generate structural excess returns versus the benchmark by avoiding fees.

KEY TAKEAWAYS

1) Replication of a trend-following benchmark index can help investors to mitigate model specification and manager selection risks which can overwhelm the benefits of allocating to the asset class.
2) However, simple regression-based replication introduces its own tracking risk driven by the trade-off between fee savings on the one hand and the inability to perfectly capture the skill of managers included in the benchmark index on the other.
3) The use of regularization techniques can help reduce skill leakage in replication, improving its tracking error to the index and allowing to more fully capture fee savings.

Keywords: *managed futures, trend-following, liquid alternatives, replication*

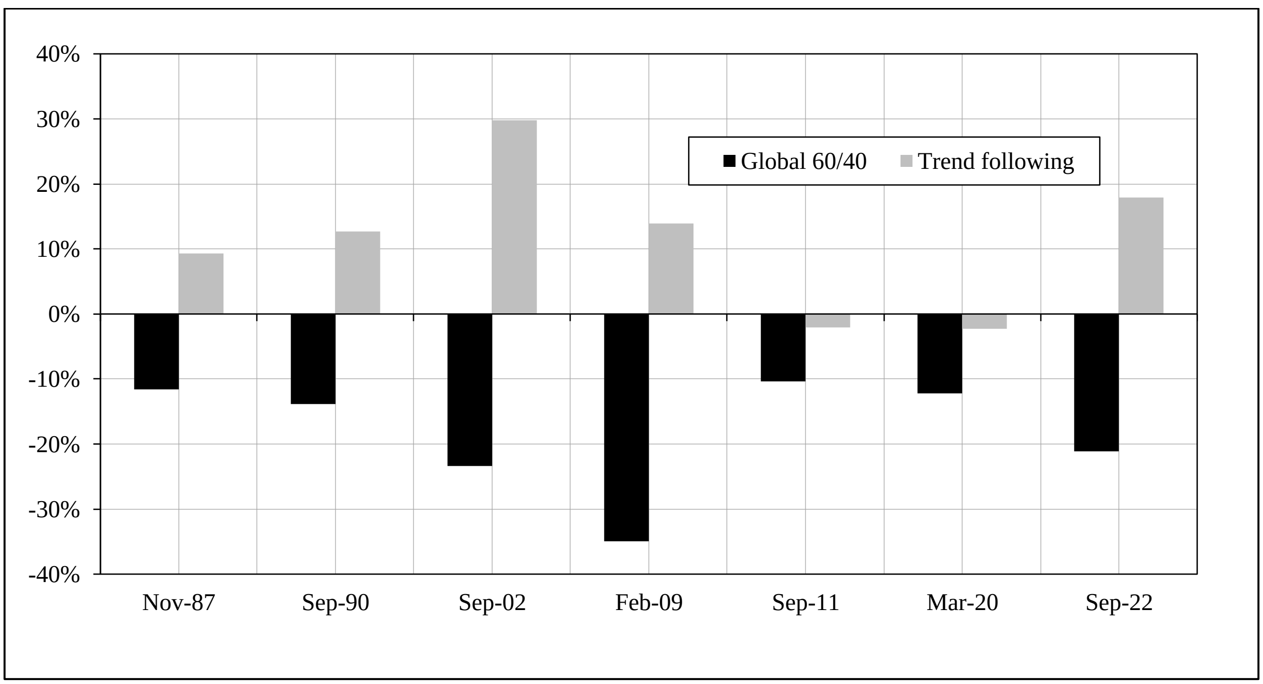
JEL Codes: *G11; G13; G15; G23*

Introduction

Trend-following in managed futures (hereafter referred to as "trend-following") is an investment strategy that aims to capitalize on asset price trends by taking long positions in rising assets and short positions in falling ones. This is typically done in the futures markets, as derivatives often provide a more efficient and cost-effective way to express leveraged investment views compared to directly using the underlying assets. The strategy has historically exhibited low correlation to stocks and bonds, as well as positive skewness, and its reliable outperformance in crisis periods has earned it the monicker "crisis alpha" (Greyserman and Kaminsky, 2014; see also Moskowitz et al. 2012; Hurst et al. 2013; Hurst et al. 2017 and Baltussen et al. 2021, who document the persistence, pervasiveness and robustness of trend-following strategies).

Indeed, the broad trend-following space (as proxied by the BarclayHedge BTOP50 Index) delivered a positive, double-digit return even in 2022 when both global equities and U.S. Treasuries registered significant drawdowns (Exhibit 1). It should, therefore, come as no surprise that even a relatively modest allocation to trend-following strategies has historically improved the return of a traditional 60/40 portfolio while reducing its volatility and maximum drawdown. Increasing strategic exposure to the space may, therefore, be an effective way to enhance portfolio resilience in the face of uncertain macroeconomic conditions and a higher interest rate regime (Maloney, 2024).

Exhibit 1: Performance of Trend-Following During Market Drawdowns



Note: Global 60/40 is a monthly rebalanced portfolio of global equities (proxied by the MSCI World Net Total Return Index) and US bonds (proxied by the Bloomberg US Aggregate Total Return Index); trend-following performance is modeled by the BarclayHedge BTOP50 Index which reflects the overall composition and style of the managed futures industry; drawdowns defined as a peak-to-trough fall in the global 60/40 portfolio exceeding 10% and cover: Aug-Nov 1987, Jan-Sep 1990, Mar 2000-Sep 2002, Nov 2007-Feb 2009, Apr-Sep 2011, Jan-Mar 2020, Jan-Sep 2022. Source: BarclayHedge and Bloomberg data.

Yet, reaping the full benefits of such strategic diversification requires transposing the general idea into a specific, implementable trading strategy. Of the many design choices that investors can make when allocating to trend-following, we focus specifically on the need to mitigate two related problems: either mis-specifying and poorly executing a trend-following program oneself or selecting a manager who does the same. Both effects – which we henceforth refer to as "manager risk" since they relate to the same underlying phenomenon – can be material enough to overwhelm the broader benefits of adding managed futures exposure to a portfolio. It is therefore natural to ask whether it might be possible to diversify across models *and* managers by replicating the performance of a broad index of trend-following funds. We tackle this question in two steps.

First, we examine the theoretical case for replication of an index of trend-following managers. We do so by designing a number of hypothetical trend-following programs which take both long and short positions across a broad range of liquid futures contracts in four key asset classes (equity indices, currencies, rates and fixed income, and commodities). We then combine the various strategies, equally weighted, into a "virtual fund of funds" ("VFOF") which can be considered a simple representation of a generic multi-manager trend portfolio. Using principal components analysis on our program's position-level return streams over the period 1992-2023 we attribute the overall performance to just a handful of macro factors. Although such analysis is backward looking, it does nonetheless suggest that – at any given moment – it should be possible to identify a relatively small basket of contracts that, when weighted appropriately, closely match the performance of our model.

Against such a background, we test the feasibility of replication both in the idealized setting of our virtual fund of funds and in a more realistic case of the BarclayHedge BTOP50 Index ("BTOP50"), which is a preeminent CTA/managed futures benchmark, comprising the largest investable managed futures programs by assets under management. While the results in both cases confirm that it is possible to replicate a large portion of index returns using simple OLS regression (a "top down" approach[2]) and just over a dozen judiciously chosen futures contracts, the resulting tracking error is not materially different from what might be expected by allocating to a small basket of CTA managers.

We thus conclude that replication comes with its own inherent model risk, yet at the same time offers the promise of structural excess returns over the benchmark through the avoidance of management and incentive fees, as well as transaction costs savings via trade netting. Still, the extent to which such "fee alpha" can be harvested is limited by the inability to perfectly capture the skill of managers included in the benchmark through the top-down process. Our numerical tests suggest that employing more complex replication models, such as regularized regression, can help reduce skill leakage, improving tracking error vis-à-vis the index and allowing to more fully capture fee savings.

[2] Top-down replication means that we are agnostic about the trend signals methodology employed and simply seek to determine optimum weights for the universe of futures such that the resulting portfolio mimics the risk/return profile of the target strategy (or a fund, or an index). Thus, top-down is different from a bottom-up replication method which seeks to identify, understand and implement the investment strategies and processes of the target fund/index.

The Challenge of Trend-Following: Manager Risk

This section sets the stage for the remainder of the paper and picks up where most trend-following research leaves off. Namely, we assume that a decision regarding the desirability of adding trend-following to a portfolio has already been made – supported by "a century of evidence" (Hurst et al., 2017) and the strategy's "crisis alpha" features (Greyserman and Kaminsky, 2014) – and focus instead on the more practical issues involved in gaining access to the space.

The first thing to note in this context is that the benefits of allocating to trend-following strategies are typically demonstrated – just as in Exhibit 1 above – using one of the broad industry indices made up of weighted returns reported by the largest funds in the category, such as the BTOP50 or the Société Générale CTA Trend Index ("SG Trend").[3] These indices, however, are not directly investable. To gain managed futures exposure, investors must either establish their own trend-following strategy or delegate the design and execution to an external manager. In either case, given the sheer number of features that can distinguish individual strategies – e.g. trend following signals, trend measurement period, portfolio construction rules, market focus, and leverage – investors' actual performance experience is likely to differ markedly from the index they originally evaluated.

To demonstrate the potential scale of such misalignment we generate 50 strategies which use the same trend measure[4] but identify trends evolving over different time periods (lookbacks) and in different investment universes (see Exhibit A.1 in the Appendix for a list of futures contracts used in the study). Specifically, we use 12 lookback periods ranging from 20 to 240 days to measure trends in 53 futures markets diversified broadly across four key asset classes: equity indices (15 contracts), currencies (8 contracts), bonds and interest rates (14 contracts), and commodities (16 contracts). Each strategy trades a randomly selected subset of markets (ranging from a minimum of 15 contracts and a maximum of 53 contracts), but all strategies trade at least the same core set of 15 markets.[5] This latter feature was incorporated to acknowledge that the largest trend followers – those who are usually included in category indices – will be forced to trade the largest and most liquid markets.

---

[3] BTOP50, an index compiled by BarclayHedge, represents the equally weighted performance of the largest trading advisor programs within the managed futures industry which together represent no less than 50% of the investable assets, and individually have at least two years of trading activity (as of 2024, there were 20 such funds in the BTOP50 Index). The majority of constituent funds are systematic, diversified and employ a range of trend-following strategies. Société Générale Trend Index is an equally weighted index of net-of-fees returns of the largest 10 managers who trade primarily in futures, are open to new investment and are recognized trend followers, exhibiting significant correlation to trend-following peers (as determined by Société Générale).

[4] The trend measure is defined as in Tzotchev et al. (2015), as $s_T = 2\Phi\left(\sqrt{T}\frac{\overline{r_T}}{\widehat{\sigma_T}}\right) - 1$, where $\overline{r_T}$ is the average daily return over the lookback period, $\widehat{\sigma_T}$ is the estimate of daily return volatility over the lookback period, and $\Phi(x)$ is the normal cumulative distribution function. This signal provides a continuous score between -1 and 1 and accounts for both the strength of the signal and its uncertainty. The signal also has the nice theoretical feature of being interpretable as the delta of a straddle with specific input parameters.

[5] The S&P 500, Euro Stoxx 50, Nikkei 225, Euro, Japanese Yen, British Pound, 10-Year US Treasury, 10-Year German Bund, 10-Year Japanese Government Bond, Light Sweet Crude, Gold, Copper, Soybeans, Corn, and Sugar No. 11.

The trading programs are designed to be always in the market, either long or short, and rebalanced daily. Individual positions are weighted to ensure risk parity[6] and scaled to a random, but constant, target portfolio volatility ranging between 10-15% annualized, which is roughly consistent with the volatility range of trend-following funds.

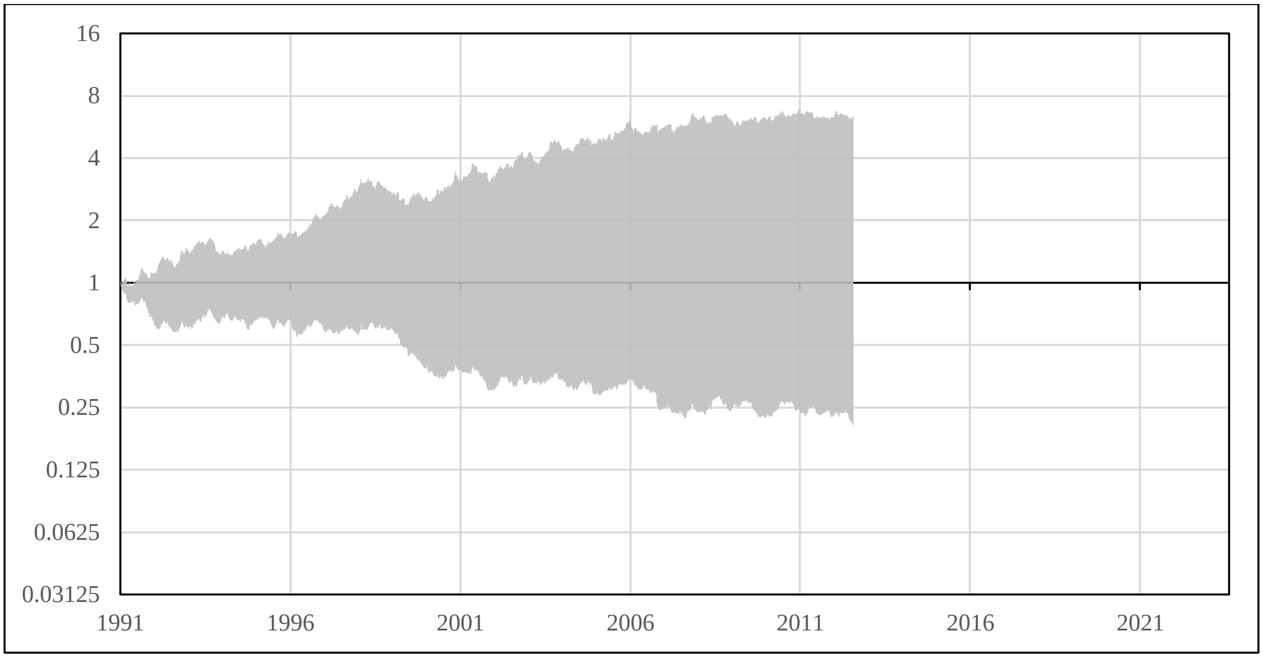
Finally, although the programs simulate trading in the largest, most liquid futures markets, the daily rebalancing can nonetheless lead to considerable turnover, especially for the quicker-moving signals. Thus, to ensure realistic comparability across strategies we subtract transaction costs, estimated as in Hurst et al. (2017), and fees made up of a flat 2% management fee and a 20% performance fee subject to a high-water mark. Our sample runs from January 1991 through end-2023, and while the availability of contracts has been uneven over time, we make sure to always have at least four futures per asset class.

Given that our strategies represent the same high-level approach, it should come as no surprise that they exhibit significant correlation of returns, averaging 0.75. However, what is striking is that seemingly small differences in specification – limited here essentially to trend lengths and contract selection – can translate into meaningfully different long-term performance. Exhibit 2 shows the range of cumulative performance delineated by the historically best and worst outcome one could hope for by following consistently one of our trend signals: the difference is roughly that of either making 7.0% or losing 8.1% per annum.[7]

Exhibit 2: Range of Cumulative Performance Outcomes for Generic Trend Strategies

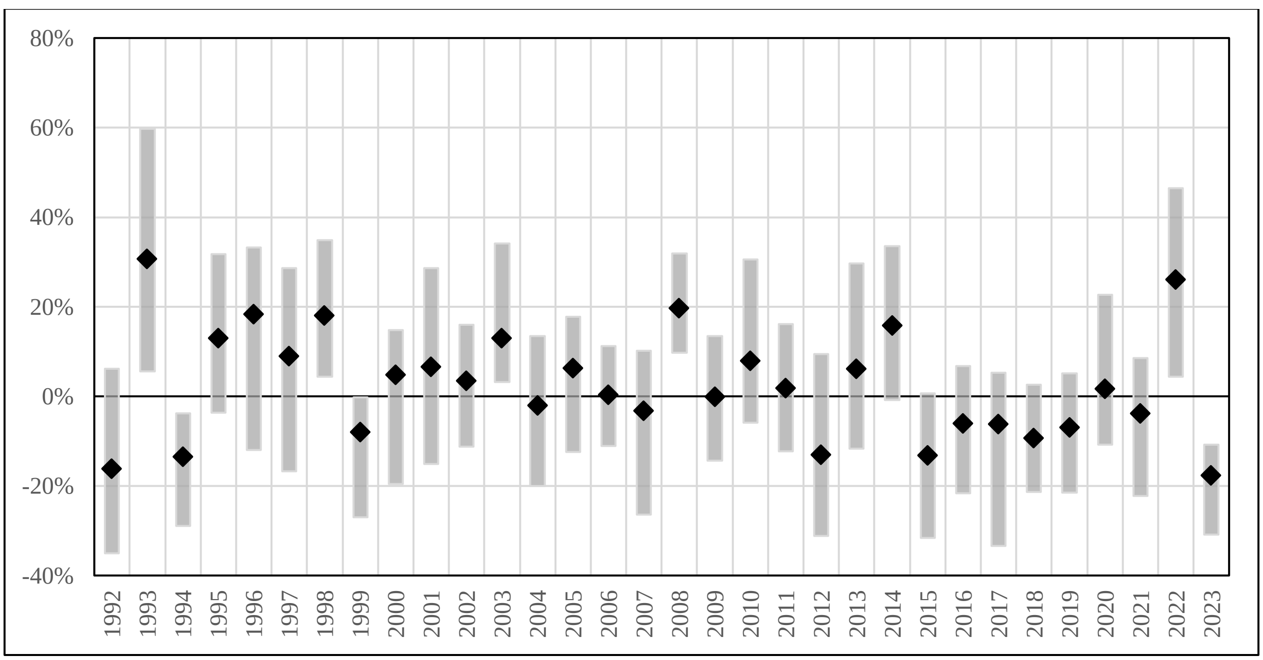
[6] The risk parity model assumes zero correlation across all markets. The risk parity algorithm seeks to find the weights such that each market contributes an equal amount of variance to the overall portfolio, with the condition that commodity sub-sectors (energies, metals, and softs) must also contribute equal variance, and that each asset class (equity indices, bonds, commodities, and currencies) contributes equal amount of variance.

[7] In-sample, the best performing model used a 240-day lookback period, and the worst performing model used a 40-day lookback period.



Note: log scale (base 2) axis applied for greater visibility; Performance is net of transaction costs and fees.

Exhibit 3: Range of Annual Excess Returns for Generic Trend-Following Strategies

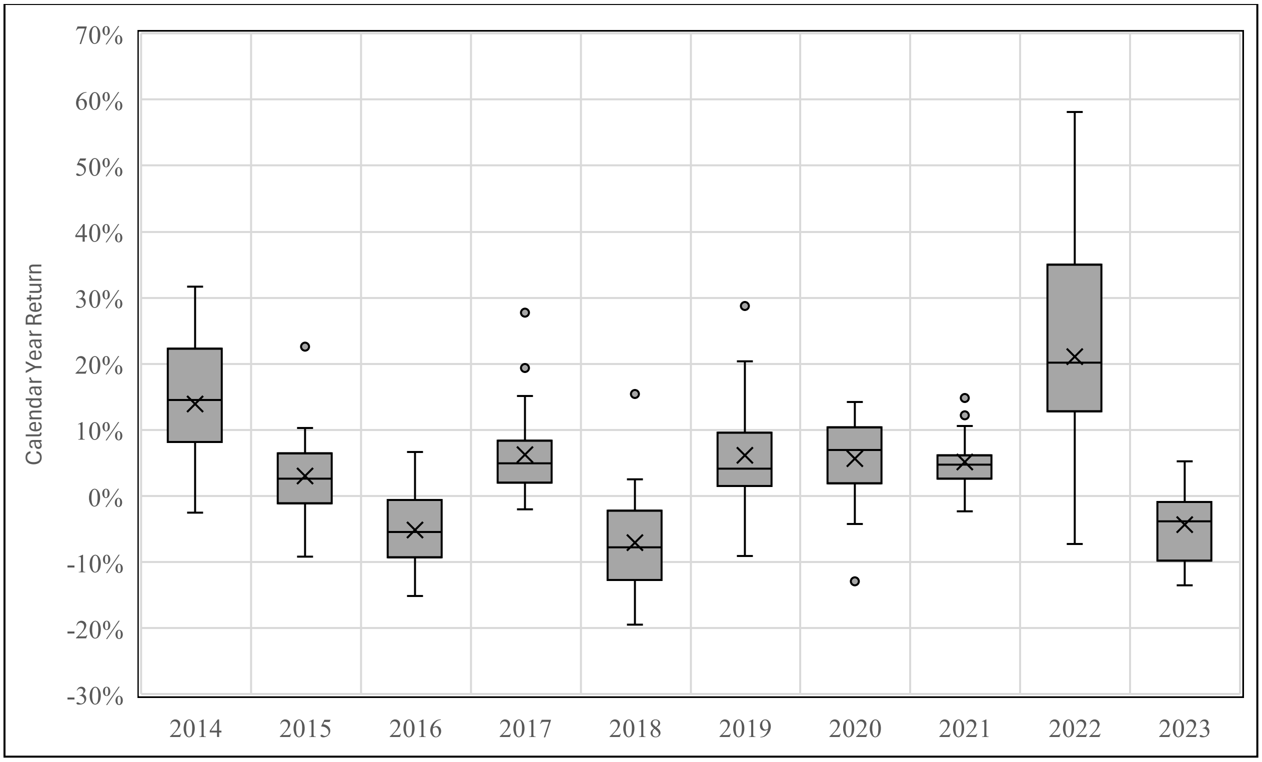


Note: the gray bars represent the dispersion of excess returns (max-min spread) for the 50 strategies, while the black diamonds represent the average excess return in a given year. Performance is net of transaction costs and fees.

Another way to appreciate the specification risk borne by following just one trend-following model is to compare the strategies' returns within the same year. Thus, Exhibit 3 shows how the models described above fared against one another over the past 32 years (1992-2023).

This is precisely an example of what we have referred to before as the manager risk associated with trend-following. That such risk is nonnegligible not only in theory but also in practice is showcased by the considerable dispersion of returns found among trend-following mutual funds over the past decade, as displayed in Exhibit 4. It should be noted that Exhibit 4 incorporates survivorship bias, using only funds that have survived into 2023. The dispersion, therefore, is likely understated, particularly with respect to relative underperformance versus the category.

Exhibit 4: Dispersion of Trend-Following Fund Returns (2014 – 2023)

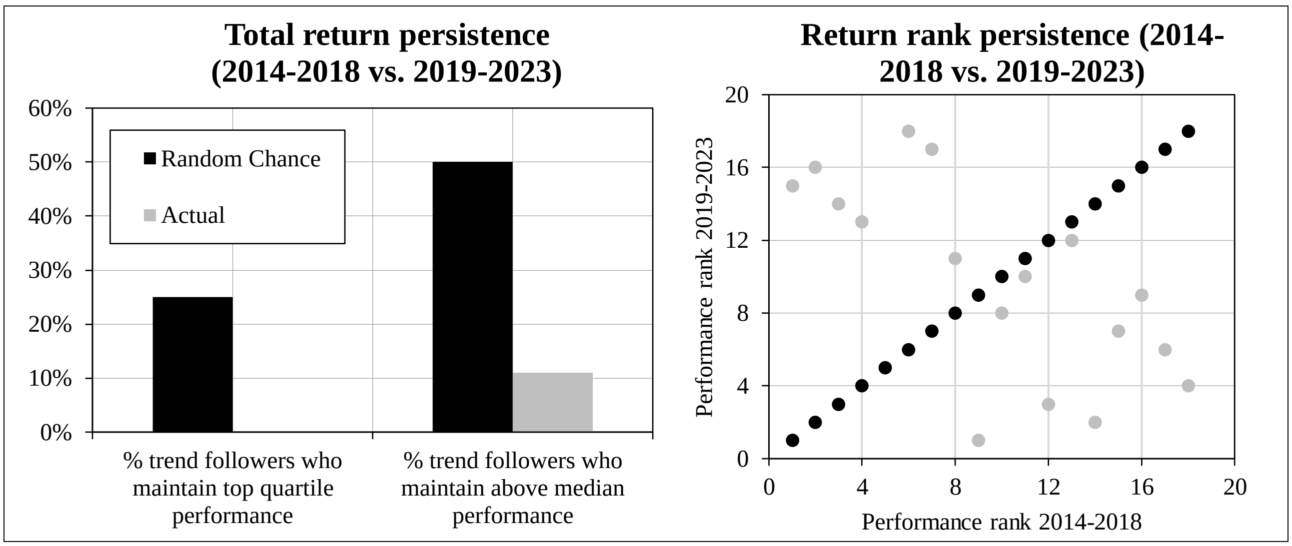


Note: the black boxes represent the interquartile range of observations for each year, the whiskers extending to the 5th and 95th percentiles; the width of the curve represents the estimated frequency of returns in a given range. The sample includes 27 largest trend following funds with a combined 2023 AuM of $27.8 billion, for which total return data is publicly available. The funds used in the analysis are identified by the following tickers: GPMFX, GPANX, QMHNX, AQMNX, GFIRX, DLMOX, RYMFX, AHLPX, GMQPX, SMFNX, PQTIX, ASFYX, SCBTA, ASDIT, ABYIX, ACXIX, AHLIX, AMFNX, CSAIX, EQCHX, EVOIX, GMSSX, LCSIX, LFMIX, LOTIX, MFTNX, RYIFX, SUPIX, WAVIX. Fund returns used in the analysis were sourced from Bloomberg.

Return dispersion among trend-following funds should not come as a surprise given the dispersion of returns documented earlier for the generic trend models, which, after all, varied essentially only in the trend length and market focus, while CTA managers can also differ in trend measures, portfolio design, risk management, execution styles, and even factor overlays, to name just a few.

However, dispersion aside, what compounds manager risk is the empirical lack of any meaningful persistence in relative fund returns, which we demonstrate in Exhibit 5 in two related ways. First, in the left-hand panel we adapt an analysis inspired by SPIVA Scorecards (Ganti et al., 2024) to show what percentage of top-quartile and above-median performers in the five-year period 2014 through 2018 managed to retain their status in the following five years. Judging by pure chance alone, we might expect 25% of the managers who delivered top-quartile returns in the first sub-period to remain in that quartile during the second sub-period. Similarly, we would randomly expect 50% of the above-median managers in 2014-2018 to again come in above median in the latter half of the decade. In fact, none of the top-quartile trend-followers in our sample, and only two of the above-median ones (11%), have managed to retain their status. Strikingly, the right-hand panel in Exhibit 5 shows that managers seem to have exhibited *negative* persistence, with the initially higher-ranking ones subsequently underperforming and the underperformers improving their ranking in the next 5 years, although the results should be treated with some caution given the small sample size and considerable changes in return dispersion between the two sub-periods.

Exhibit 5: Performance Persistence in Trend-Following Fund Returns



Note: the original sample is the same as in Exhibit 4, but for purposes of this analysis funds with less than 10 full years of performance data through December 2023 were excluded, bringing the total population to 18.

It could be argued that investors seeking to diversify single-manger risk should spread their bets by allocating to a couple, say 3-4 funds, or even opt for a fund of funds or multi-manager structure. While these approaches can help mitigate single-manager risk, they introduce operational hurdles and additional costs, particularly with private funds – a point we shall return to below. For now, suffice it to say that allocating among several funds requires ongoing manager due diligence, creates friction in rebalancing, eliminates the opportunity to reduce transaction costs through trade netting, and introduces the "basket of options" problem with performance fees.[8]

[8] The "basket of options" problem is when an allocator pays performance fees to underlying managers even though their aggregate portfolio sits below its high-water mark. It is called "basket of options" because performance fees are akin to selling a call option to the manager, and in effect the allocator has sold calls on their underlying positions rather than their aggregate portfolio. This is, effectively, a dispersion trade and the performance difference will depend upon the relative volatility and correlation of underlying managers.

Examining The Case for Replication

To solve the manager selection problem, we consider an alternative solution: the replication of a category benchmark index. By design, the method we propose seeks to gain asset-class-level exposure with minimal tracking error, thus helping to bridge the gap between the expected and delivered portfolio contribution of the trend-following theme.

That replication can work in a space marked by such large variation in how managers express their strategies may not be readily apparent. To illustrate this, we return to the 50 trend-following strategies introduced in the prior section and combine them, equally weighted, into a single diversified program. Such a "virtual fund-of-funds" ("VFOF") can be considered a simple representation of a multi-manager trend portfolio and allows us to peer behind the curtain of trend-following strategies and better understand their performance drivers.

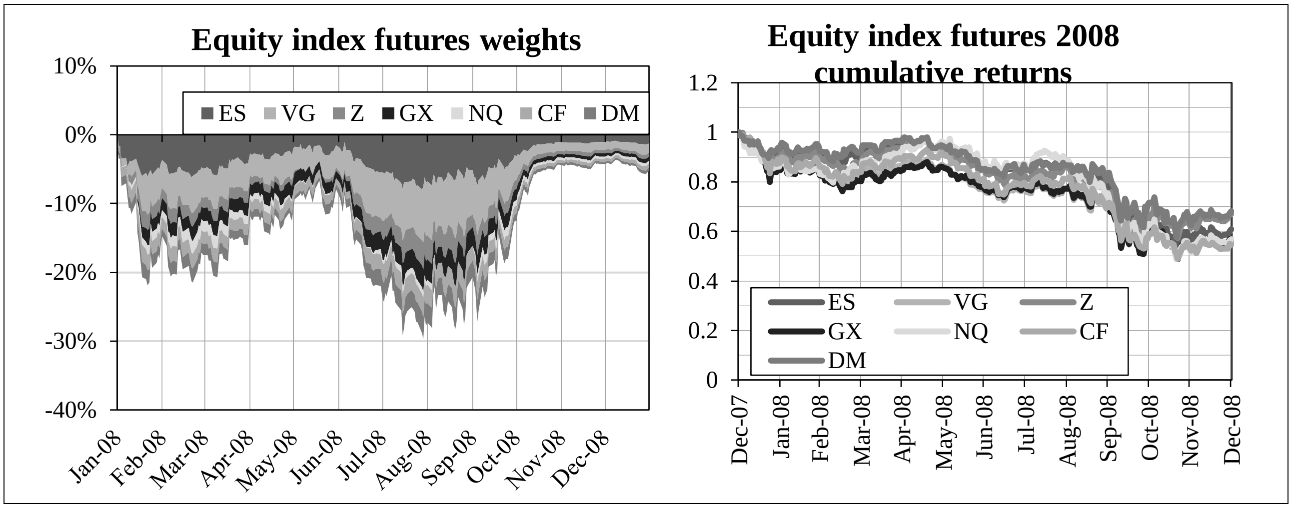
The opportunity in replication is "fee alpha." For our hypothetical VFOF model, over the in-sample period we estimate an annualized cost of 200 basis points ("bps") for manager fees, 110bps for performance fees, and 275bps for unnetted transaction costs. [9,10] This amounts to 585bps of potential annualized cost savings.

Recall first that since the underlying strategies are always either long or short in the 53 futures contracts, in periods of heightened correlation, VFOF performance may result from running what effectively amounts to the same trade implemented in different markets. A good illustration of that effect is provided by equity instruments. Exhibit 6 shows the allocation weights of the VFOF to seven major equity index futures next to their annual cumulative gross returns. The volatility-scaled positions taken in equity indices were very similar throughout the year, as were the return paths for the seven contracts. In hindsight, we know that as the crisis began to unfold, correlations increased, and different idiosyncratic factors were dominated by a single systematic one. Hence, at least in 2008, trading just one of these instruments would have likely sufficed to generate the bulk of the return of the entire asset class. What about other markets and years?

Exhibit 6: VFOF Equity Allocations and Cumulative Returns in 2008

[9] We measure the potential in-sample annualized savings by calculating versions of the VFOF with and without performance fees, measuring their performance difference.

[10] Given modern trading costs, we estimate approximately 3bps of savings per 100 percentage points of one-way turnover avoided through trade netting. To generate our estimate, we calculate a version of the VFOF where no-trade netting occurs between managers and a version where trades are first netted before transaction costs are applied.



Note: Virtual fund of funds allocates equally to all 36 trend-following models introduced in Exhibit 2. Futures symbols represent: ES – S&P 500, VG – Euro Stoxx 50, Z – FTSE 100, GX – DAX, NQ – Nasdaq 100, CF – CAC 40, DM – Dow Jones. Source: Bloomberg data.

To answer quantitatively, we rely on principal components analysis, inspired by a recent exposition by Quantica (2024). We run the analysis not on contract returns themselves but on the time series of position-level returns (i.e. position-weighted futures returns). The eigenvectors of the resulting covariance matrix will be the statistical risk factors whose returns, judiciously weighted,[11] will add up to the fund's gross daily return, thus providing the desired performance attribution to independent factors. Since our futures universe contains 53 different contracts, we will eventually end up with 53 different contributing factors. However, using the concept of Shannon's entropy applied to normalized absolute return contributions, we can further calculate the *effective* number of independent risk factors driving our P&L (in a sense, calculate the diversity index for our contributing factors).[12]

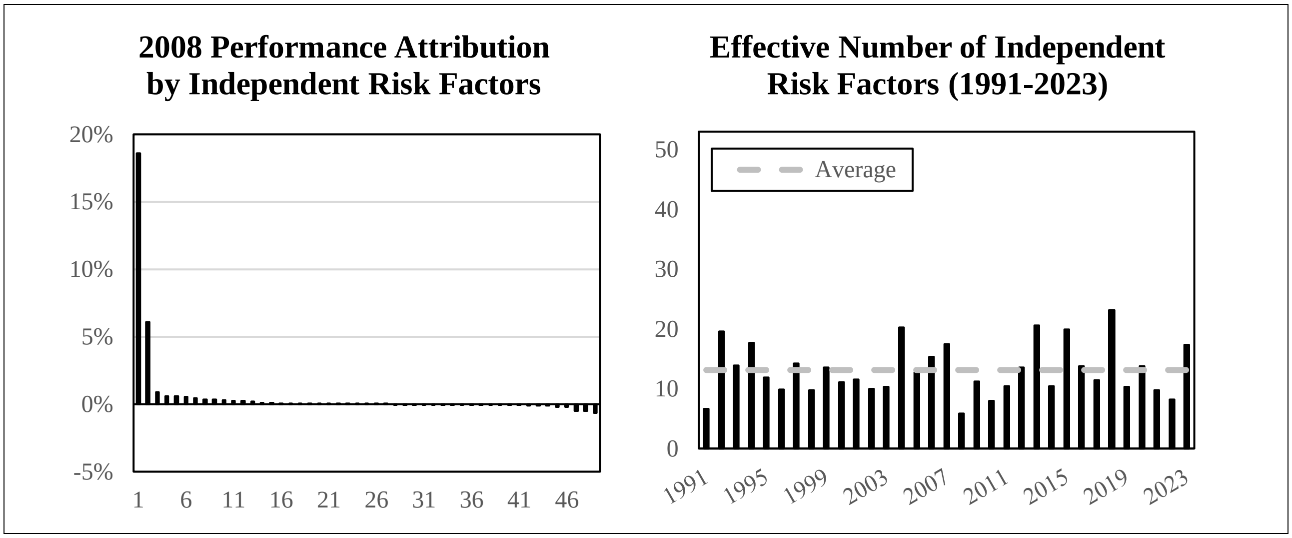
Exhibit 7 shows an application of this technique to the 2008 return of our fund of funds (left-hand panel) and a calculation of the effective number of independent factors required to explain performance of the fund over the entire sample (right-hand panel). Note first that just as we have intuited above, although the VFOF allocated to 53 different contracts, there were just a handful of independent factors that truly mattered, and one factor that contributed over two-thirds of the total return. Nor is the 2008 experience unique: while the right-hand panel of Exhibit 7 makes clear that the effective number of factors fluctuated somewhat in the past decade, the

[11] Formally, given the eigenvectors matrix E, we find implied factor weights W, by solving the linear system: $E \times W = [1, \ldots, 1]^T$; the unit vector on the right-hand side represents 100% invested by the fund in each of the futures markets.

[12] This part of our analysis is inspired by a recent Quantica (2024) report which calculates the number of effective factors as the exponential of the Shannon entropy. Formally, letting $\tilde{c}_i = |c_i| / \sum_i |c_i|$ be the normalized absolute return contribution associated with i-th principal component (c_i being the non-normalized contribution, i=1,…,53), the effective number of independent factors is given by: $\exp\left(-\sum_i \tilde{c}_i \log \tilde{c}_i\right)$. The effective number of factors ranges from 1 (when there is only a single factor contributing 100% of the return) to 53 if all 53 factors contribute equally.

return streams are far from perfectly diversified, with a long-term average number of effective factors equal to 13.1 and just 5.6 and 8.1 in crisis periods like 2008 and 2022 respectively. If these purely statistical factors converge towards macro-economic features (e.g. the most important feature in 2022 could be described generically as "short bonds / long USD"), it may be possible to replicate the VFOF with just a limited subset of the contract universe, particularly if we believe those will be the most important macroeconomic factors going forward.

Exhibit 7: VFOF Factor-Based Attribution



Note: risk factors estimated by running principal components analysis on virtual fund's position returns and deriving implied factor weights using the eigenvectors matrix so that return contributions of all factors add up to the fund's return in a given year; Effective number of factors (right-hand panel) is calculated as the exponential of the Shannon entropy of the normalized absolute factor return contributions (see footnote 6 for details); the effective number of independent factors ranges from 1 (no diversification) to the number of contracts used by the fund, here 53 (full diversification).

VFOF Replication: A Best-Case Scenario

To test whether index replication could work as suggested by the PCA analysis, we start with a best-case, top-down approach. Specifically, we try to find the set of markets that minimizes the full-period in-sample tracking error of the top-down approach (our approach is outlined in Exhibit 8). Note that this is, explicitly, an exercise in overfitting designed to determine a best-case outcome.[13] Exhibit 9 zooms in on the results plotting tracking errors for the optimal ticker sets across different set sizes.

Exhibit 8: VFOF Leave-One-Out Replication

[13] Strictly speaking, the estimated tracking errors are subject to some path dependency, and hence only a conditional minimum; the leave-one-out analysis does not explore all possible paths to find the global minimum tracking error as a function of the number of contracts. Furthermore, the contracts once discarded are not reevaluated in the context of different ticker sets. Thus, a hypothetical portfolio build-up approach which would start with a single contract and iteratively add more futures might result in a different, potentially lower, tracking error profile than the leave-one-out analysis which moved from the full universe down to one contract.

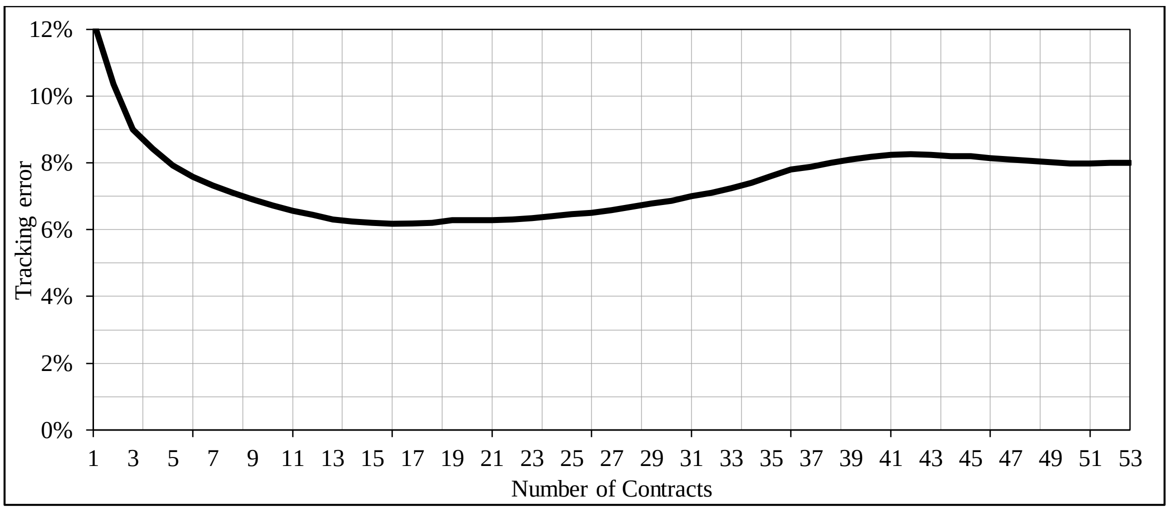
- Step 1: Perform top-down replication of VFOF returns regressing them on the full set of contracts (i.e. 53) in a rolling 40-day window; as before, individual positions are weighted to ensure risk parity and scaled to a target volatility of 12.5% annualized;

- Step 2: Smooth the estimated regression coefficients ("betas") and shift them by two days to avoid look-ahead bias before using them to construct an index of fitted VFOF returns;

- Step 3: Calculate the full-sample tracking error between the fitted index and VFOF;

- Step 4: Leave-one-out loop: after calculating the tracking error for the full set of tickers (n=53), test removing one contract at a time from the current set of tickers (n) and repeat Steps 1-3 for the remaining (n-1) tickers, record tracking errors and discard the worst performing futures (i.e. one whose removal leads to the lowest tracking error deterioration); continue iteratively, with one ticker being discarded in each loop until only a single ticker remains.

Exhibit 9: Top-Down Replication of VFOF: Tracking Error by Number of Contracts



Note: Replication based on "leave-one-out" analysis as described in Exhibit 8; VFOF returns are net of fees and estimated transaction costs; tracking error is annualized, and the sample runs 1992-2023; The dashed line depicts the minimum tracking error attained of 6.2% at 16 contracts utilized.

We find that tracking error ranges from almost 12% when using just one contract to around 6% when 16 contracts are used. Interestingly, the number of contracts maximizing the fit is just above the number of effective factors we estimated above. This may suggest that the macro-economic features driving trend-following returns may be well captured by a simple set of futures markets. Though, it is worth acknowledging that the trough in Exhibit 9 may simply be governed by numerical stability in the top-down regression, driven by the quantity of data (i.e. our 40-day lookback) versus the number of explanatory variables.[14]

[14] To alleviate this concern, we performed robustness tests performing a similar exercise using regularized regression techniques (lasso, ridge and elastic net) with cross-validation to mitigate the risk of overfitting. The results were not substantially different and are available from the authors upon request. In light of this, to facilitate presentation, we limit ourselves below to the simplest case of ordinary least squares regression.

It is important to stress that these results are meant to capture a *best-case* result, and the tracking errors reflect an in-sample lower bound, which would not necessarily generalize to future periods. Still, these lower bounds show that even without any prior knowledge about the trend-following signals used within the VFOF, top-down replication was able to achieve a daily return correlation to the VFOF of 0.86 with just 16 futures markets.

Still, with a 6.2% annualized tracking error – achieved, effectively, with a crystal ball – it is fair to say that replication introduces its own unique model risk, and we are essentially swapping one form of manager risk for another. Indeed, the tracking error for replication falls just below both the mean and the median tracking error of underlying managers (6.6% and 6.4% respectively).[15]

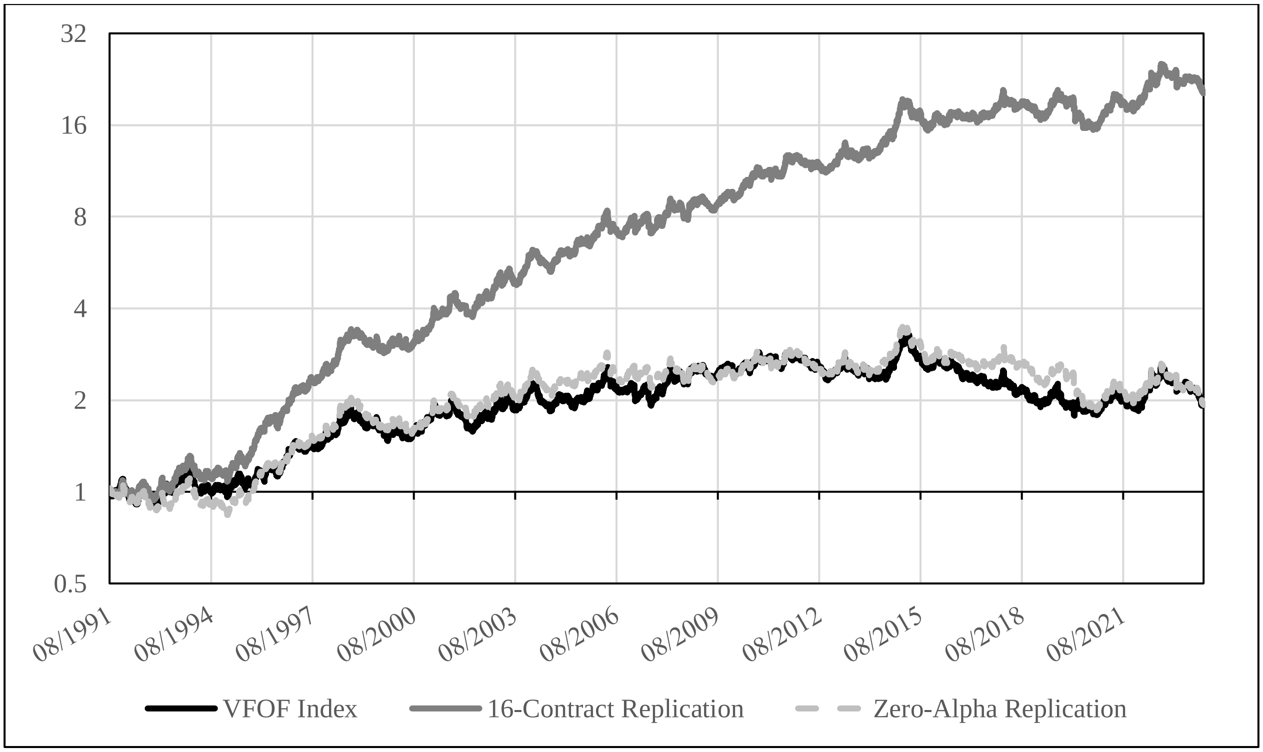
While selecting a small basket of managers at random would have achieved a lower tracking error on average, random selection does not come with any expectation of generating excess returns versus the benchmark.[16] Nor does it solve the selection problem regarding which managers will necessarily exhibit high or low tracking error going forward. A replication approach, on the other hand, may be able to generate excess returns with respect to the benchmark by avoiding manager fees, performance fees, and transaction costs (through trade netting).

As mentioned previously, the opportunity for replication is approximately 585bp of annualized fee savings. With a tracking error of 620bps, one might argue that this hypothetical replication model is an active strategy with an information ratio of 0.94. In sample, the 16-contract model identified above exceeded these estimates, generating an annualized return of 9.8% net of estimated transaction costs, versus 2.1% for the VFOF. In Exhibit 10, we plot the cumulative excess returns for the VFOF, the 16-contract model, and the 16-contract model assuming zero excess returns versus the VFOF to demonstrate the quality of fit.

Exhibit 10: Cumulative Returns for the VFOF Index and the 16-Contracts Replication Model

[15] With average correlation of ρ=0.87 to the VFOF and a volatility σ=12.5%, the expected tracking error of a single manager is $TE = \sigma\sqrt{2(1-\rho)} \approx 6.4\%$.

[16] This is not strictly true, as increasing the number of managers will impact the "basket of options" problem discussed in footnote 7.



Note: log scale (base 2) axis applied for greater visibility; VFOF index returns are net of fees and estimated transaction costs; 16-Contract Replication returns are net of estimated transaction costs; Zero-Alpha Replication is based on the estimated replication portfolio but assumes zero excess returns versus the VFOF; the sample runs August 1991 – December 2023.

While instructive and useful in building intuition, it should be stressed again that our VFOF replication is in many ways a best-case scenario: managers have no skill, charge high management and performance fees, trade a fixed and known universe, trade at precisely the same time of day and pay identical transaction costs. On top of that, we use hindsight bias to find an ex-post best fit replication model. Thus, there is little to no opportunity for "skill leakage" relative to the fee alpha we can capture. To better estimate the trade-off between fee alpha and skill leakage in practice, we need to deploy our replication framework in a more realistic setting of a composite of live managed futures trend-following funds.

Replicating the BarclayHedge BTOP50 Index

A reasonable candidate for testing replication in a more realistic setting is the BTOP50 index ("BTOP50"), a closely tracked benchmark in the managed futures space comprising the largest investable managed futures programs by assets under management. As the BTOP50 reflects the performance of actual managers, we expect greater dispersion in methodologies, variation in execution timing, efforts to reduce transaction costs, and a general expression of craftsmanship alpha. Furthermore, based upon the current members of the BTOP50,[17] we

[17] 2024 constituents of the BTOP50 include AlphaSimplex Group (Managed Futures), AQR Capital (MF Moderate Vol), Aspect (Diversified Fund), Campbell & Co (CMF), CIBC (Active Currency Overlay 12%), Crabel (Gemini 1x – M), DBi Managed Futures, Graham Capital Management (Tactical Trend A), John Street Capital (Trident), Lynx

estimate an average annual management fee of 1.2% and an average performance fee of 12.6% (with 25% of members not charging a performance fee), reducing the opportunity for fee alpha.

In seeking to replicate the BTOP50, we first apply a naïve OLS regression, leveraging insights gleaned from the previous section that reasonably accurate replication may be accomplished with a relatively small number of contracts. In practice, it is sensible to assume that our selected universe should require exposures from each of the major asset classes and should focus on the most liquid and prominent contracts (as they are more likely to be traded by the majority of underlying funds).

In that vein, we select a universe of fifteen contracts: three each in equity indices, fixed income and currencies, plus six commodity contracts. The specific contracts are chosen based on a combination of liquidity[18], economic exposure, and geographic exposure to ensure that the resulting set holds a reasonable probability of being representative of the universe traded by the underlying managers. Since liquidity and other characteristics might change from year to year, we try to limit potential lookahead bias by considering the selection criteria as of January 2010, which is when our replication exercise begins. The constituents of this universe can be found in Exhibit 11.

Exhibit 11: 15-Contract Replication Universe

Category	Ticker	Name
	ES	E-Mini S&P 500
Equities	VG	Euro Stoxx 50
	NK	Nikkei 225
	EC	Euro
Currencies	JY	Japanese Yen
	BP	British Pound
	TY	US 10-Year
Bonds	RX	German 10-Year
	JB	Japanese Government Bonds
Energies	CL	Light Sweet Crude
Precious Metals	GC	Gold
Industrials	HG	Copper
	S	Soybeans
Softs	C	Corn
	SB	Sugar No. 11

Using this universe, the naïve top-down approach generates an annualized tracking error of 4.4%. As with the previous section, this level of tracking error roughly corresponds to the

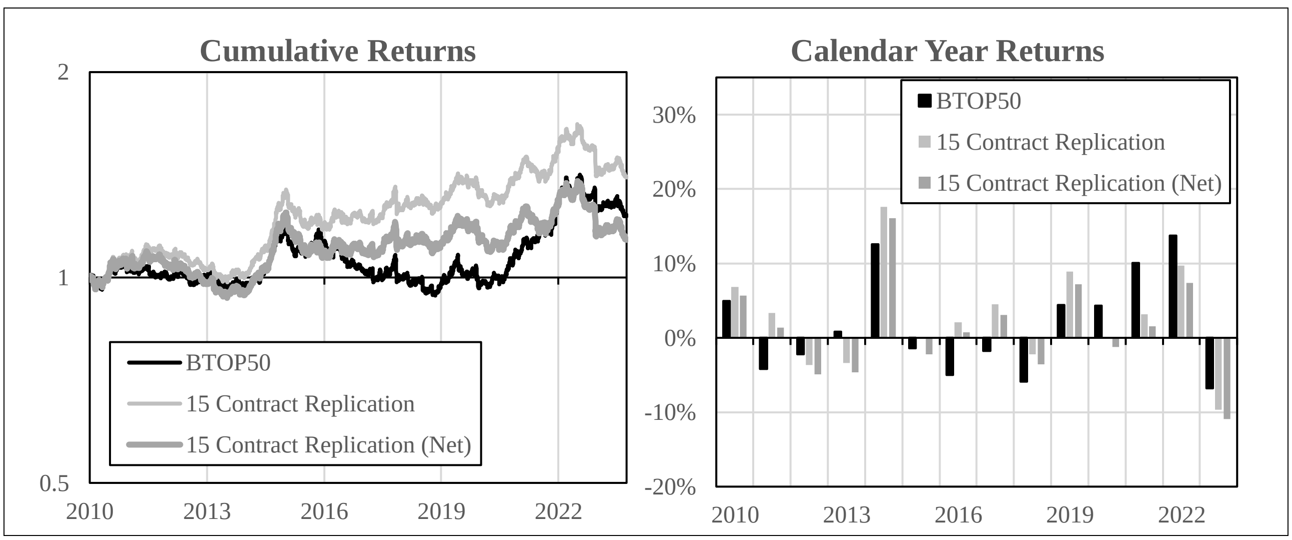
Asset Management (Lynx Program USD), Man (AHL Alpha), Millburn Ridgefield (Diversified), PIMCO (Trends), Polar Star Management SEZC (Polar Star Ltd), Quest (AlphaQuest Original – AQO), ROW Asset Management (Diversified), Statar Capital (Natural Gas), Systematica (BlueTrend Fund A), Transtrend (DTP/Enhanced Risk – USD), and Winton Capital Management (Div).

[18] In the selection of the commodity contracts for each of the universes, the contracts were selected in relation to the size of the contracts of the DowJones-UBS Commodity Index (now known as the Bloomberg Commodity Index or BCOM) from January 2010 (published in October 2009). As this index selects the constituent weights based on liquidity metrics and economic production, this provides a proxy for liquidity and potential economic importance to be used for the selection criteria. Additionally, as these constituent weights were available prior to the start date of the analysis, this reflects information that would have been known at the time, in the effort of minimizing hindsight bias.

tracking error of individual fund managers. Indeed, the tracking error of the managers analyzed in Exhibit 4 ranged from 3.6% to 14.6% with an average of 7.4%. Managers' higher tracking error is, to some extent, a byproduct of the higher average volatility of their strategies.[19] Therefore, to ensure a more accurate comparison, we scale fund returns to match the volatility of the BTOP50 which reveals that the tracking error of the replication portfolios lies squarely between that of the average volatility-scaled fund (5.1%) and the bottom quartile result (4.0%).

While our naïve top-down replication approach does not outperform in terms of tracking accuracy, it comes with the potential of generating consistent excess returns versus the benchmark by avoiding fees. Over the full period, our 15-contract replication outperformed the BTOP50 by 99bp annualized and realized an information ratio of 0.23 (Exhibit 12). While this information ratio falls short of our expectations (an information ratio of 0.27 is theoretically achievable by simply avoiding management fees, to speak nothing of performance fees or transaction costs), it is worth noting that with only 13.5 years of data, the confidence interval around the estimate is quite wide.

Exhibit 12: Cumulative Returns (left panel) and Calendar Year Returns (right panel) for the BTOP50 Index and the 15-Contract Replication Index



Note: 15 Contract Replication returns are net of estimated transaction costs; 15 Contract Replication (Net) returns are net of transaction costs, management fees, and incentive fees[20]; and the sample runs 2010-2023; in panel 1, 2010 is a partial year with the sample beginning in April; BTOP50 Index performance is net of the 1-3 month T-bill rate.

More importantly, as already hinted above, we should acknowledge the possibility that managers within the BTOP50 are, in aggregate, performing with skill that cannot be replicated by our process. For example, they may be generating significant profit from a long tail of alternative

---

[19] Recall that tracking error scales linearly with volatility; see footnote 15.

[20] Based on information provided by BarclayHedge at the time of writing, the average management fee charged by the constituents of the BTOP50 Index is 1.2%, while the average incentive fee is 12.6%. The 15 Contract Replication (Net) returns are net of these management and incentive fees, in addition to transaction costs. See footnote 19 for a list of the constituents included in the index.

asset classes whose performance cannot be proxied by our 15-contract set. Or they may be incorporating signals or processes that operate at a speed that cannot be captured by our 40-day lookback (e.g. stops, breakouts, or mean-reversion signals).

In seeking to improve the naïve top-down replication framework, we focus on two potential sources of skill leakage: universe selection and signal speed.

To allow for a larger universe of contracts, we resort to LASSO, which works on the entire set of available regressors but uses regularization to enforce sparsity.[21] By allowing us to work with the entire universe, but varying the relevance of specific contracts over time, this approach should not only help us to more closely mimic the return profile of BTOP50 managers, but also address any risk of hindsight bias in our selection of the 15-contract universe.

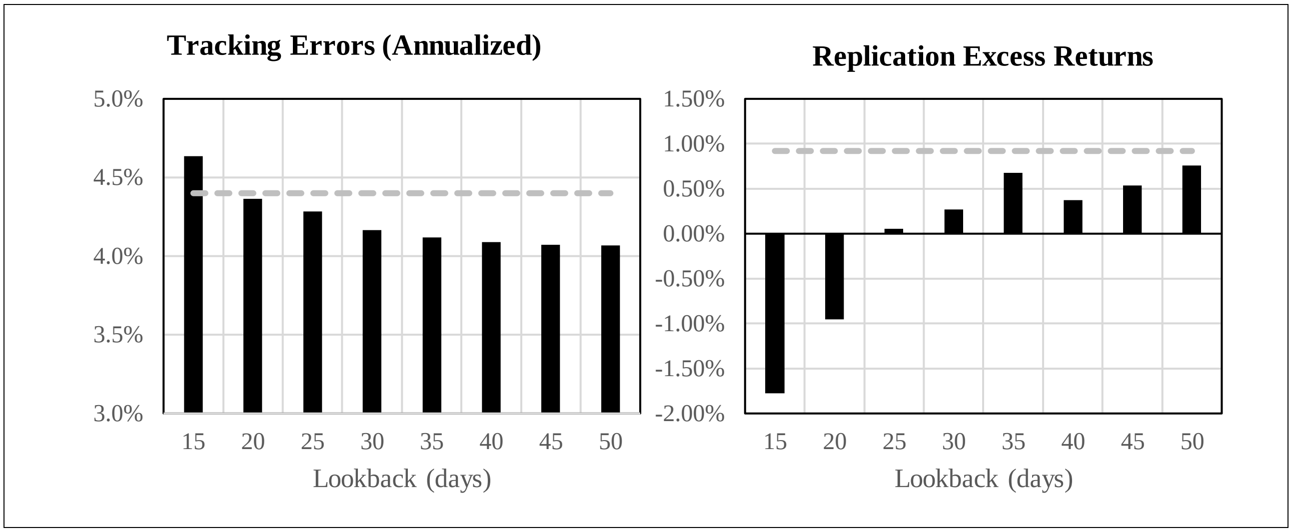
Exhibit 13 shows the tracking errors of replication portfolios estimated using LASSO on all 53 contracts over different rolling windows (ranging from just 15 to 50 days) as well as their excess returns to the benchmark. It also plots the tracking error and excess return realized by the OLS method as a reference point. Although there is some improvement in tracking accuracy relative to OLS for all but the shortest lookback, excess returns decrease.

While it is challenging to assess the significance of these differences given that they come from a sample spanning only 13.5 years, one potential factor behind the subpar performance of LASSO portfolios is the costs of trading a larger contract set. Indeed, comparing the 40-day LASSO portfolio to its OLS counterpart on a gross return basis reveals that regularization improved tracking error by 40 basis points annualized, while trailing the simpler method by 55 basis points after transaction costs. This suggests further gains may be achieved by means of reducing turnover, such as weight smoothing or trading size thresholds.

Here we may be able to solve multiple shortcomings: by combining the results of different replication lookback windows, we may be able to capture trends of various lengths, reduce noise trades through model averaging, and improve cost savings through trade netting. In this vein, we construct a replicating portfolio by averaging the contract weights estimated by our LASSO models over 5 different lookback periods: 30, 35, 40, 45, and 50 days.

Exhibit 13: LASSO vs. OLS Replication Tracking Errors and Excess Returns to BTOP50

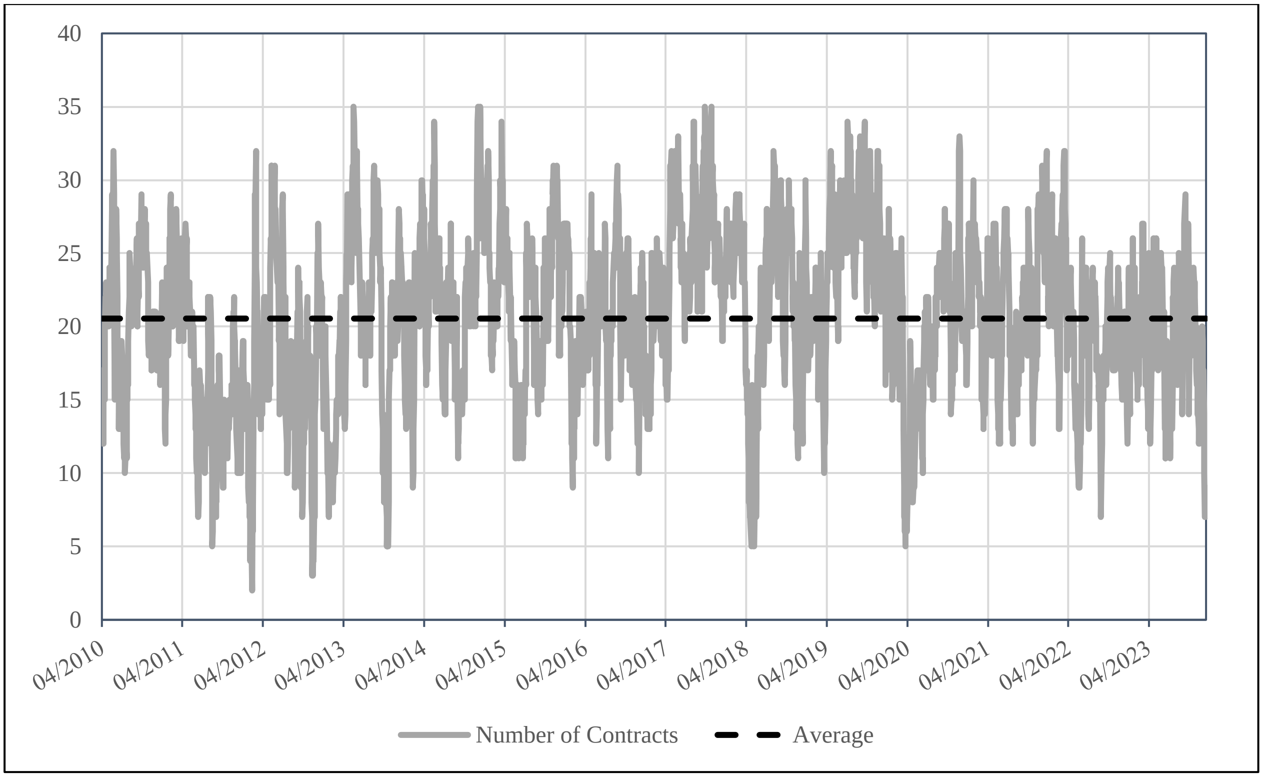
[21] Formally, while OLS minimizes $(y - X\beta)'(y - X\beta)$, LASSO (Least Absolute Shrinkage and Selection Operator) minimizes $(y - X\beta)'(y - X\beta) + \alpha|\beta|$, where α is a penalty (regularization) term which – loosely speaking – controls how much we care about keeping coefficients small versus fitting the data well. Our implementation of LASSO automatically selects the optimal regularization strength for each lookback window, not only making estimation possible (where OLS would not be feasible) but also ensuring greater stability of parameters and out-of-sample fit of our replication.



Note: tracking errors and excess returns are relative to the BTOP50 index; the dashed horizontal line represents the tracking error and excess return of the naïve OLS regression, respectively.

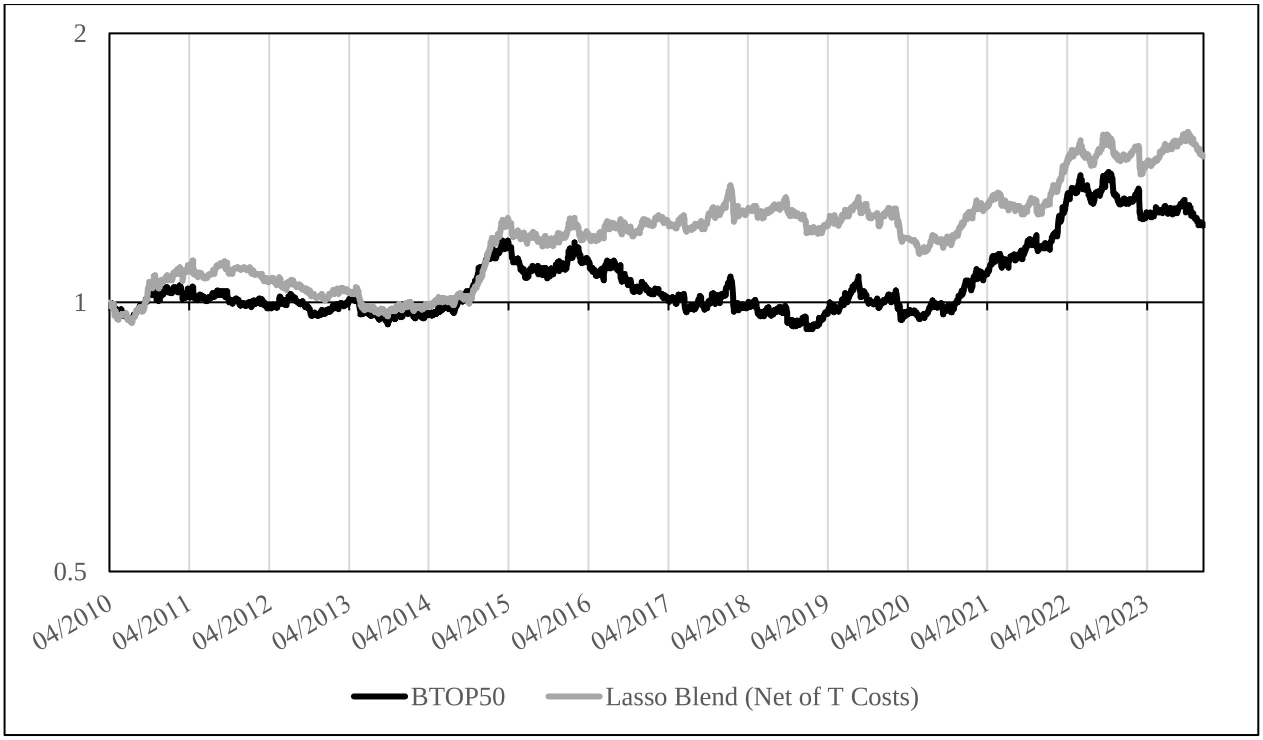
Exhibit 14 displays the number of holdings over time in such a blended portfolio. We find, as we have already stressed above, it is neither prudent nor in most cases necessary to use a very broad universe of contracts in replication. Even though we run our model with all 53 contracts, regularization prevents it from too aggressively fitting noise, so that the resulting replicating portfolio holds on average only about 20 contracts – i.e. just 5 more than the naïve approach – with the total count fluctuating between as few as 3 and as many as 35. Importantly, *which* contracts are selected varies significantly over time, with each of the 53 contracts being selected in at least 10% of periods.

Exhibit 14: Number of Markets Traded in LASSO Replication Model Over Time



Combined with the positive effect of averaging weights from different lookbacks, this approach not only reduces the tracking error of the blended replication portfolio to below 4%, but also improves its returns (net of transaction costs), thereby increasing the in-sample information ratio of the regularized ensemble strategy to 0.34 from 0.23 for the simple OLS replication.

Exhibit 15: Cumulative Excess Return for the BTOP50 Index and the LASSO Blend Replication Model



Note: Lasso Blend represents the performance (net of transaction costs) of a portfolio constructed by averaging the weights from top-down LASSO replication models estimated using lookback periods of 30, 35, 40, 45, and 50 days.

Conclusion

Throughout this article, we have stressed that for all the virtues of trend-following, implementation involves the risk of either mis-specifying and poorly executing a trend-following program oneself or selecting a manager who does the same, with potentially grave consequences. We believe index replication provides a viable solution to this conundrum by allowing investors to diversify across both models and managers without the operational burden and costs this would otherwise entail. The replication approach developed – centered around a parsimonious set of liquid futures contracts and a simple regression model – demonstrated an ability to closely track the performance of the BarclayHedge BTOP50 Index, generating consistent excess returns through avoidance of management fees, performance fees, and trade netting.

The top-down regression model employed has several attractive features. First, it is agnostic to the methods employed by underlying managers. Second, it has the flexibility to adapt to methodology drift over time. Finally, it has the potential to capture the performance of trading assets not held in the replication universe (including synthetic assets such as spreads) if such assets can be approximated by a linear combination of in-universe holdings.

The method is not without its drawbacks, however. For example, numerical stability is a concern when choosing a tradeoff between the number of assets and the regression period. Furthermore, the rolling regression approach will inherently lag sudden changes in manager positioning. Additionally, when an underlying asset exhibits little contributed variance over the lookback period, the regression methodology will not be able to determine whether underlying managers hold long, flat, or short positions. Finally, the top-down process can struggle to fully capture the skill with which underlying managers identify and monetize trends.

We believe these shortfalls can be addressed, however, using regularization techniques, additional parameterizations, or by introducing complementary replication approaches (e.g. bottom-up). Although all the above can help to some extent, any replication strategy will likely retain some residual tracking error. Hence, we argue, replication should not be viewed as a panacea for single manager risk, but as an active strategy which keeps that risk at an acceptable level, while generating excess returns versus the benchmark through the avoidance of fees and transaction costs savings.

References

Baltussen, G., Swinkels, L., & Van Vliet, P. (2021). Global factor premiums. *Journal of Financial Economics*, 142(3), 1128-1154.

Blitz, D. (2022). Expected Stock Returns When Interest Rates Are Low. *Journal of Portfolio Management*, 48(7).

Federal Reserve Bank of New York. (2024). *Responses to survey of primary dealers*. Markets Group, July 2024.

Ganti, A.R., Nelesen, J., Di Gioia, D. & Longo S. (2024) *U.S. Persistence Scorecard*. S&P Dow Jones Indices, May.

Greyserman, A., & Kaminski, K. (2014). *Trend following with managed futures: The search for crisis alpha*. John Wiley & Sons.

Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2013). Demystifying managed futures. *Journal of Investment Management*, 11(3), 42-58.

Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2017). A Century of Evidence on Trend-Following Investing. *Journal of Portfolio Management*, 44(1), 15.

Israel, R., Jiang, S., & Ross, A. (2017). Craftsmanship alpha: An application to style investing. *The Journal of Portfolio Management*, 44(2), 23-39.

Maloney, T. (2024). Honey, the Fed Shrunk the Equity Premium: Asset Allocation in a Higher-Rate World. *Journal of Portfolio Management*, 50(6).

Moskowitz, T. J., Ooi, Y. H., & Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, 104(2), 228-250.

Quantica. (2024). *Trend-following and risk factor diversification in 2022 and 2023: a tale of two extremes*. Quarterly Insights, No. 17, March.

Tzotchev, D., Kolanovic, M., Lau, A., Krishnamachari, R.T., and Silvestrini, D. (2015). *Designing robust trend-following system: Behind the scenes of trend-following*. J.P. Morgan Research Paper.

Appendix A

Exhibit A.1: Futures summary statistics						
Category	**Ticker**	**Name**	**Start Date**	**CAGR**	**Annualized Volatility**	**Max Drawdown**
EQUITIES	ES	E-Mini S&P 500	10.09.1997	5.7%	19.4%	-62.9%
EQUITIES	DM	Dow Jones	05.04.2002	6.7%	18.1%	-54.3%
EQUITIES	NK	Nikkei 225	05.09.1988	0.4%	24.0%	-82.0%
EQUITIES	NQ	Nasdaq 100	22.06.1999	7.1%	26.8%	-84.7%
EQUITIES	Z	FTSE	29.02.1988	2.5%	20.0%	-65.7%
EQUITIES	VG	Euro Stoxx 50	04.01.1999	2.4%	25.0%	-65.6%
EQUITIES	HI	Hang Seng	02.04.1992	4.1%	26.5%	-66.8%
EQUITIES	GX	DAX	26.11.1990	4.4%	23.5%	-71.9%
EQUITIES	CF	CAC 40	08.12.1988	10.6%	40.3%	-63.0%
EQUITIES	XP	SPI 200	03.05.2000	5.1%	22.6%	-68.7%
EQUITIES	IB	IBEX 35	01.07.1992	4.0%	24.7%	-63.2%
EQUITIES	KM	KOSPI	06.05.1996	1.3%	35.3%	-87.5%
EQUITIES	QC	Swedish OMX	15.02.2005	6.7%	27.6%	-67.9%
EQUITIES	ST	MIB	23.03.2004	3.0%	26.1%	-73.5%
EQUITIES	RTY	R2K	11.07.2017	5.0%	24.8%	-43.3%
CURRENCIES	JY	Japanese Yen	23.05.1986	-1.9%	10.5%	-74.8%
CURRENCIES	EC	Euro	20.05.1998	-0.9%	9.2%	-46.7%
CURRENCIES	BP	British Pound	28.05.1986	0.7%	9.7%	-50.2%
CURRENCIES	SF	Swiss Franc	07.04.1986	0.6%	11.1%	-50.8%
CURRENCIES	AD	Australian Dollar	13.01.1987	2.1%	11.4%	-42.3%
CURRENCIES	NV	New Zealand Dollar	08.05.1997	1.6%	12.5%	-42.6%
CURRENCIES	CD	Canadian Dollar	04.04.1986	0.6%	7.4%	-35.3%
CURRENCIES	PE	Mexican Peso	26.04.1995	4.6%	11.6%	-39.3%
BONDS	TU	US 2-Year	26.06.1990	1.0%	1.6%	-8.7%
BONDS	FV	US 5-Year	23.05.1988	2.1%	3.9%	-17.5%
BONDS	TY	US 10-Year	31.12.1985	3.0%	6.2%	-23.9%
BONDS	WX	Ultra US 10-Year	24.11.2009	0.9%	3.4%	-19.4%
BONDS	JB	Japanese Government Bonds	31.12.1985	2.7%	4.3%	-18.3%

BONDS	DU	German 2-Year	10.03.1997	0.2%	7.4%	-35.0%
BONDS	OE	German 5-Year	07.10.1991	2.1%	3.3%	-14.5%
BONDS	RX	German 10-Year	26.11.1990	2.6%	11.0%	-37.6%
BONDS	ED	Eurodollar	02.04.1986	0.4%	0.8%	-3.4%
BONDS	ER	Euribor	09.12.1998	0.1%	0.4%	-2.6%
BONDS	IR	AUD 90-day	13.06.1989	0.4%	0.9%	-3.4%
BONDS	WB	US Treasury Bonds	24.11.2009	0.1%	1.6%	-10.5%
BONDS	WN	Ultra Long-Bond	12.01.2010	1.9%	14.3%	-52.7%
BONDS	XM	Australian 10-Year	21.09.1987	-0.4%	6.7%	-45.0%
ENERGIES	CL	Light Sweet Crude	31.12.1985	3.2%	39.0%	-108.3%
ENERGIES	XB	RBOB	04.10.2005	7.7%	39.7%	-82.3%
ENERGIES	HO	Heating Oil	01.07.1986	9.9%	35.2%	-88.9%
ENERGIES	NG	Natural Gas	04.04.1990	-15.7%	51.7%	-99.9%
METALS	GC	Gold	31.12.1985	1.5%	15.7%	-73.4%
METALS	SI	Silver	31.12.1985	-0.2%	28.5%	-82.4%
METALS	HG	Copper	07.12.1988	3.9%	25.4%	-68.8%
SOFTS	LH	Lean Hogs	02.04.1986	-1.6%	25.0%	-94.1%
SOFTS	LC	Live Cattle	31.12.1985	2.6%	14.6%	-49.3%
SOFTS	C	Corn	31.12.1985	-5.0%	24.1%	-95.5%
SOFTS	W	Wheat	31.12.1985	-7.5%	27.7%	-98.0%
SOFTS	CT	Cotton No. 2	31.12.1985	0.4%	25.9%	-93.0%
SOFTS	S	Soybeans	31.12.1985	4.2%	21.6%	-73.1%
SOFTS	SB	Sugar No. 11	31.12.1985	1.8%	32.9%	-83.9%
SOFTS	KC	Coffee	31.12.1985	-6.7%	35.4%	-97.0%
SOFTS	CC	Cocoa	31.12.1985	-1.8%	28.4%	-93.6%