Portfolio Construction Using Nonlinear Polymodels

RAPHAEL DOUADY

CNRS, University Paris 1-Pantheon-Sorbonne

AUGUST 20-24, 2022

FUNDAÇÃO GETULIO VARGAS (FGV), RIO DE JANEIRO, BRAZIL

Decision Making in Finance

- What is the Upside?
  - Why should I invest?
  - Is there a path to success?
  - Is the median scenario profitable?
  - Is there the possibility of winning the lottery?
- What is the Downside?
  - Distribution of returns: wrong answer!!!
  - Describe scenarios, with all their consequences
  - Are losses limited?
  - What are the risk sources? What can trigger risky scenarios?

Manager Selection and Categorization

Will *forthcoming* performance be good?

- Was it good in the past? $\Rightarrow$ Will it remain?
- Will *forthcoming* environment be favorable to the strategy?

If *past* performance was good

- Was it Skill?
- Was it Luck?
- Was it Contingent to the Environment? $\Rightarrow$ Will it remain?
  - *Beta*, not *Alpha*

In case it turns bad, will also turn bad for *all* the portfolio?

- Are there killing scenarios? Which ones? How likely?
- Can I hedge the risks?

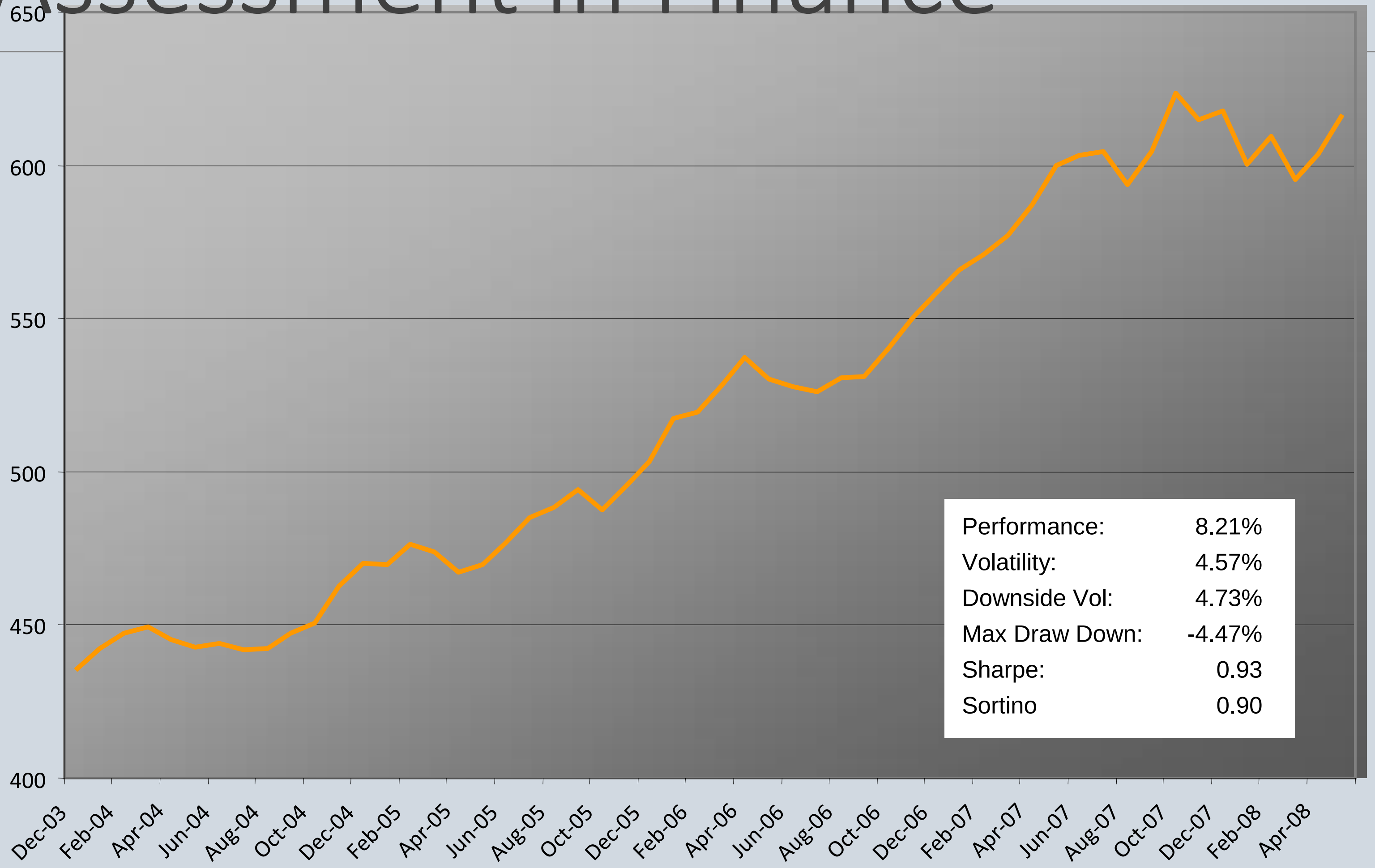
Risk Assessment in Finance

- What are the sources of risk? Which scenarios are dangerous?
  - Which variables X_i have an influence on Y <u>when it moves dangerously</u>?
  - How far can they go?
- How far can it go? How frequently? How much should I pay (or charge) for an insurance?
  - What is the shape of the return distribution?
  - Fat tails?
- Can I anticipate dangerous events? Can I protect myself against?
  - Crisis prediction (Type I and Type II errors)
  - Hedge efficiency (nature of the hedge, insurer's quality, liquidity risk…)
- Diversification
  - Portfolio return distribution
  - Contribution of each component to the overall risk
  - Portfolio optimization

Risk Assessment in Finance

This fund seems to display all possible green lights for an investor…

But will the performance last?

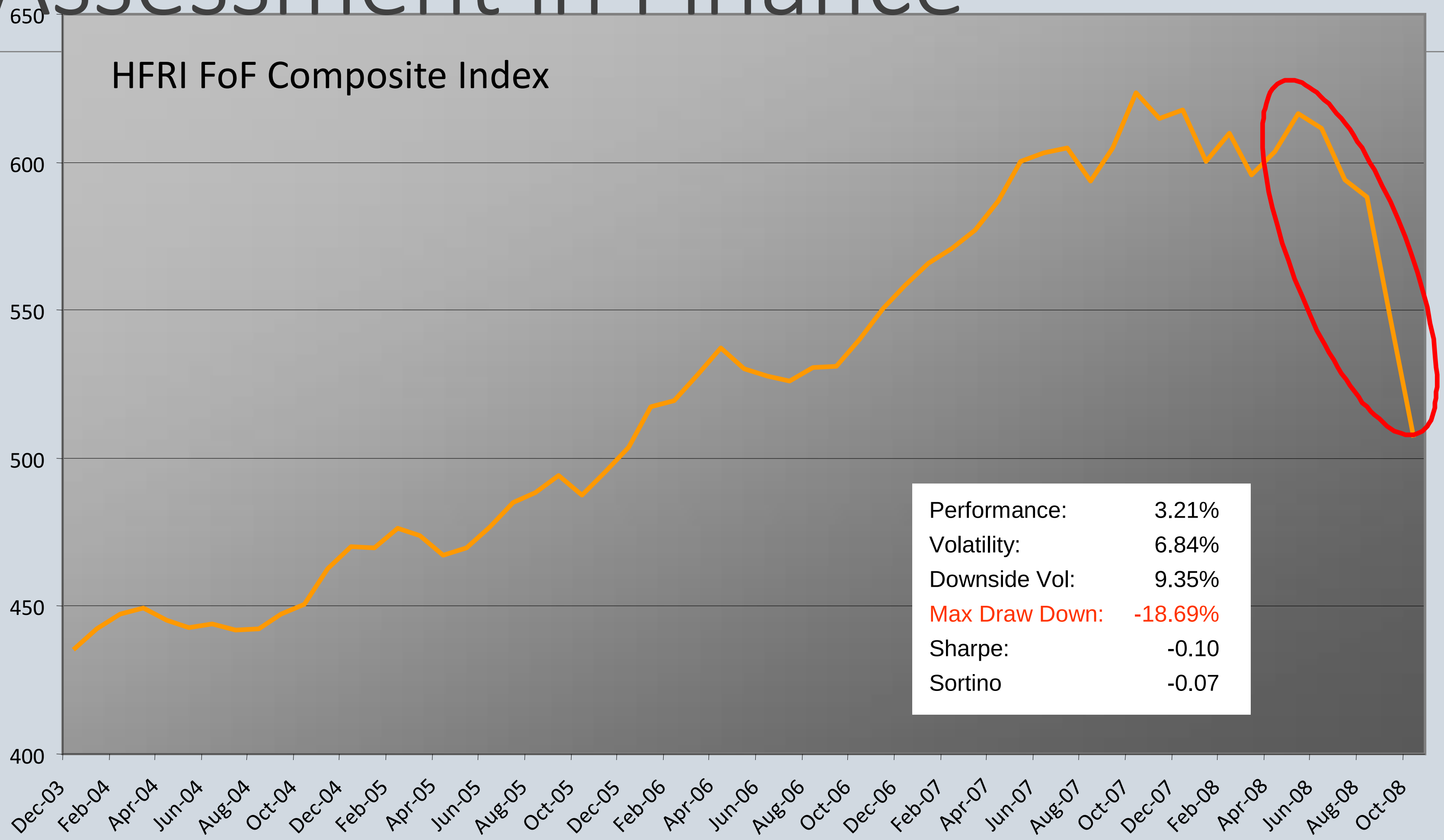


Risk Assessment in Finance

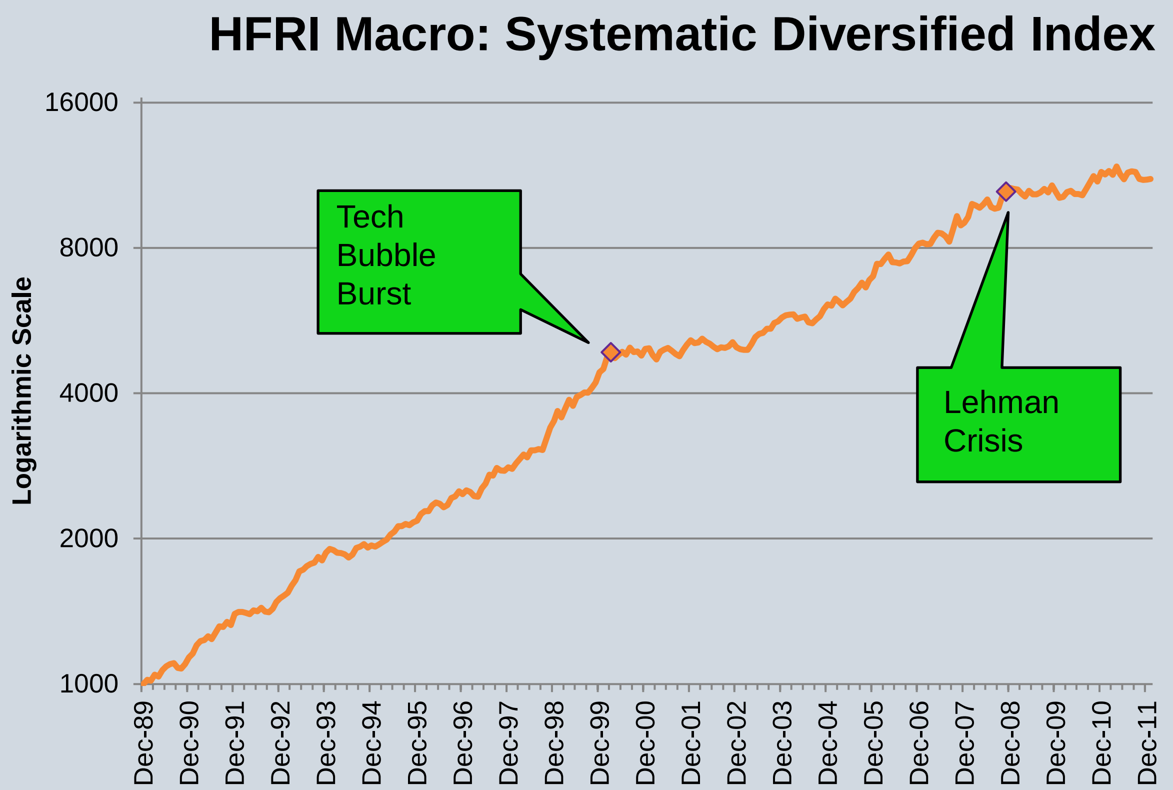
NO! Losses during the crisis exceeded 4 times the previous Max Drawdown…

The fund?

The HFR Fund of Funds index!



CTA Index History



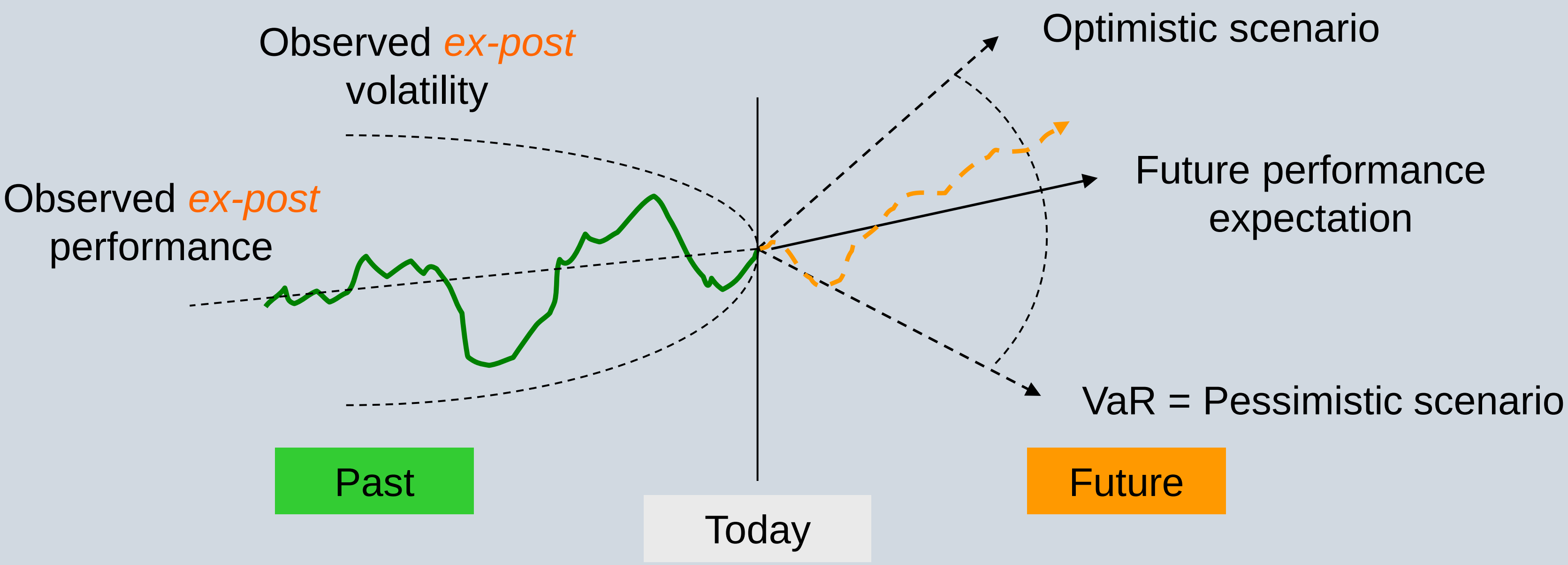
Trend Following stagnates in post-crisis regime

DOUADY – AI AND INVESTMENT

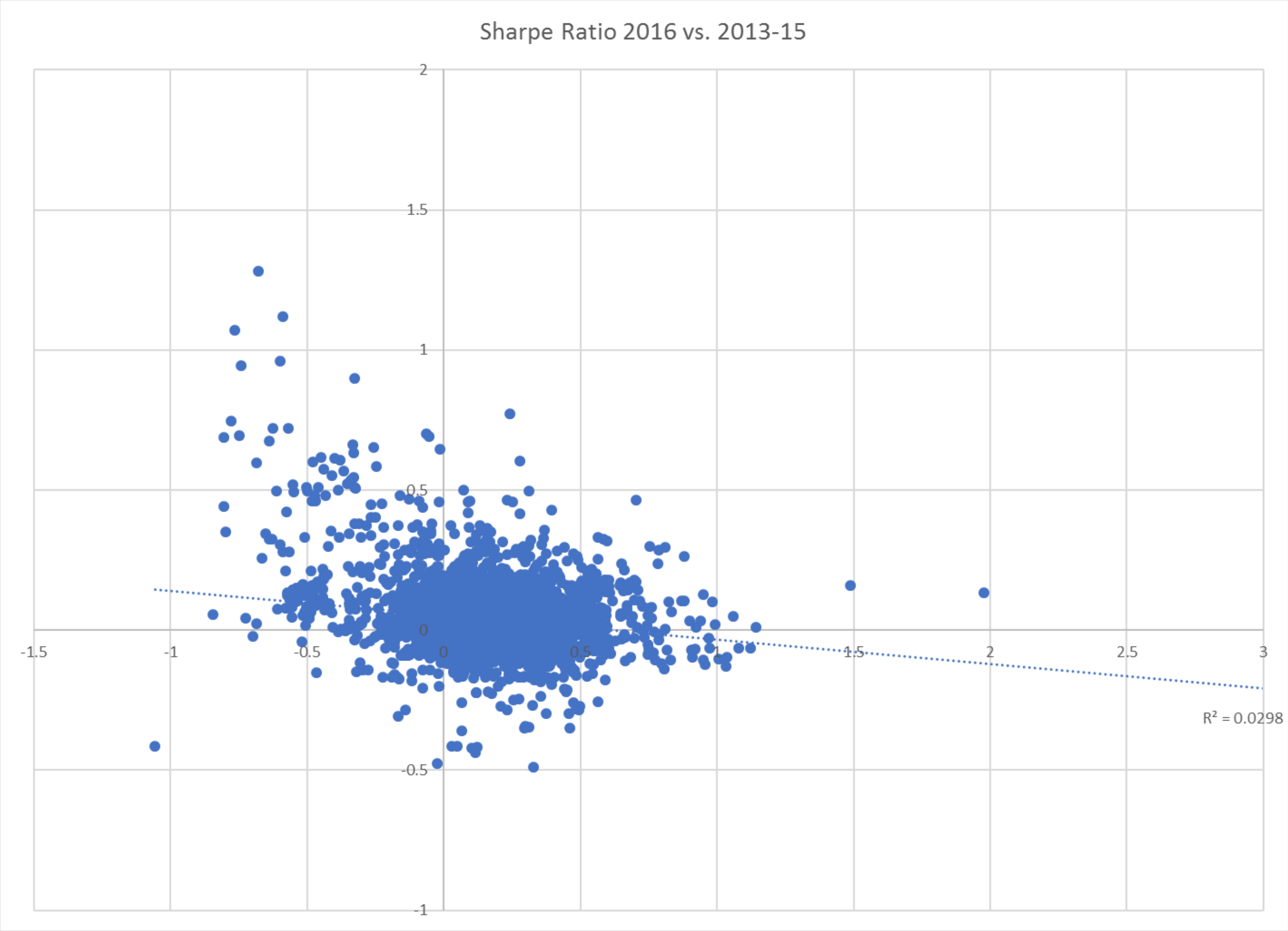
What is a Risk Measure?

A *Risk Measure* is an *Ex-ante* measure

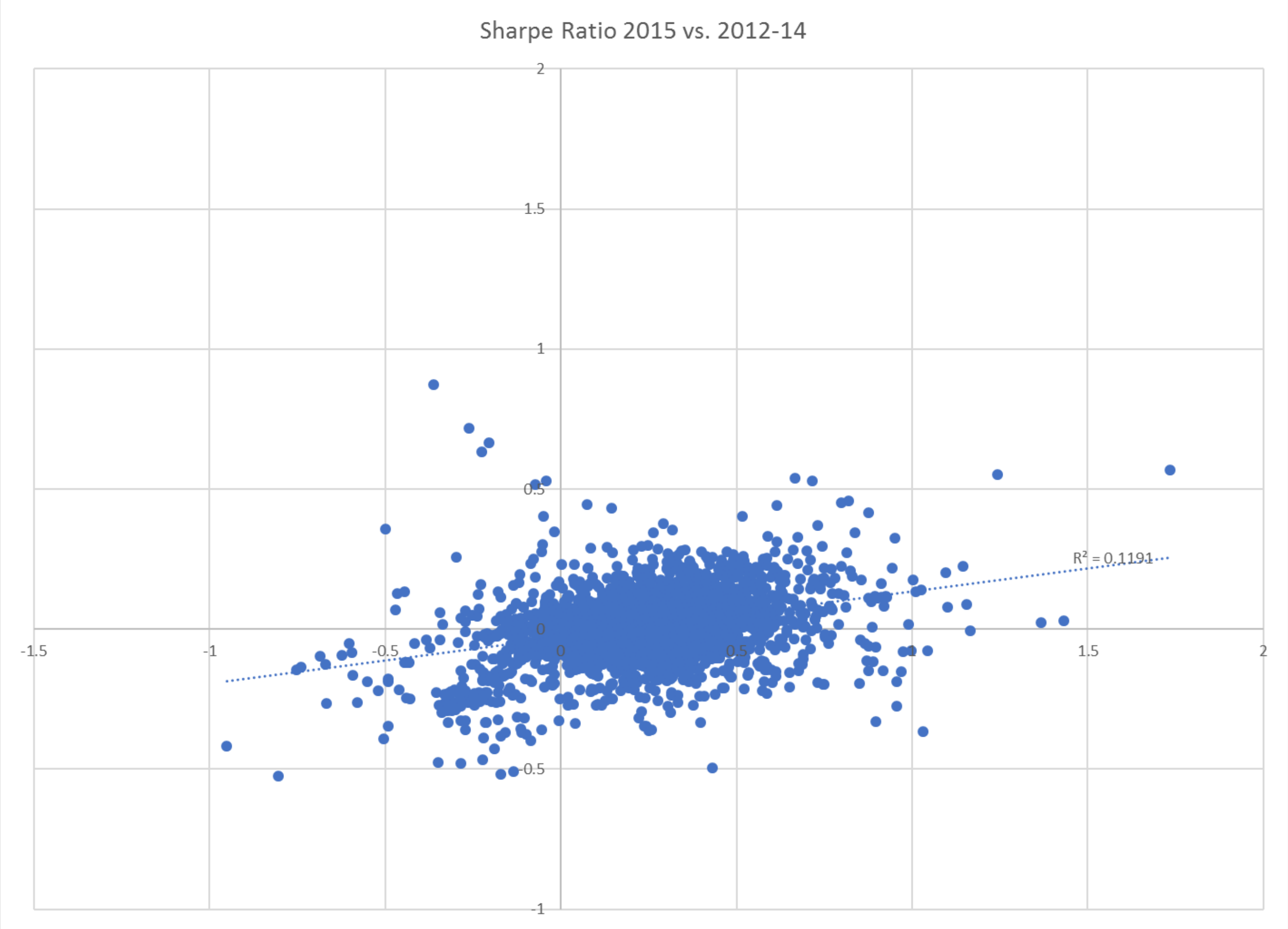
- **Question 1:** *"What is the Range of possible Future returns?"*
- **Answers:** Expectations, Value-at-Risk, *ex-ante* Volatility



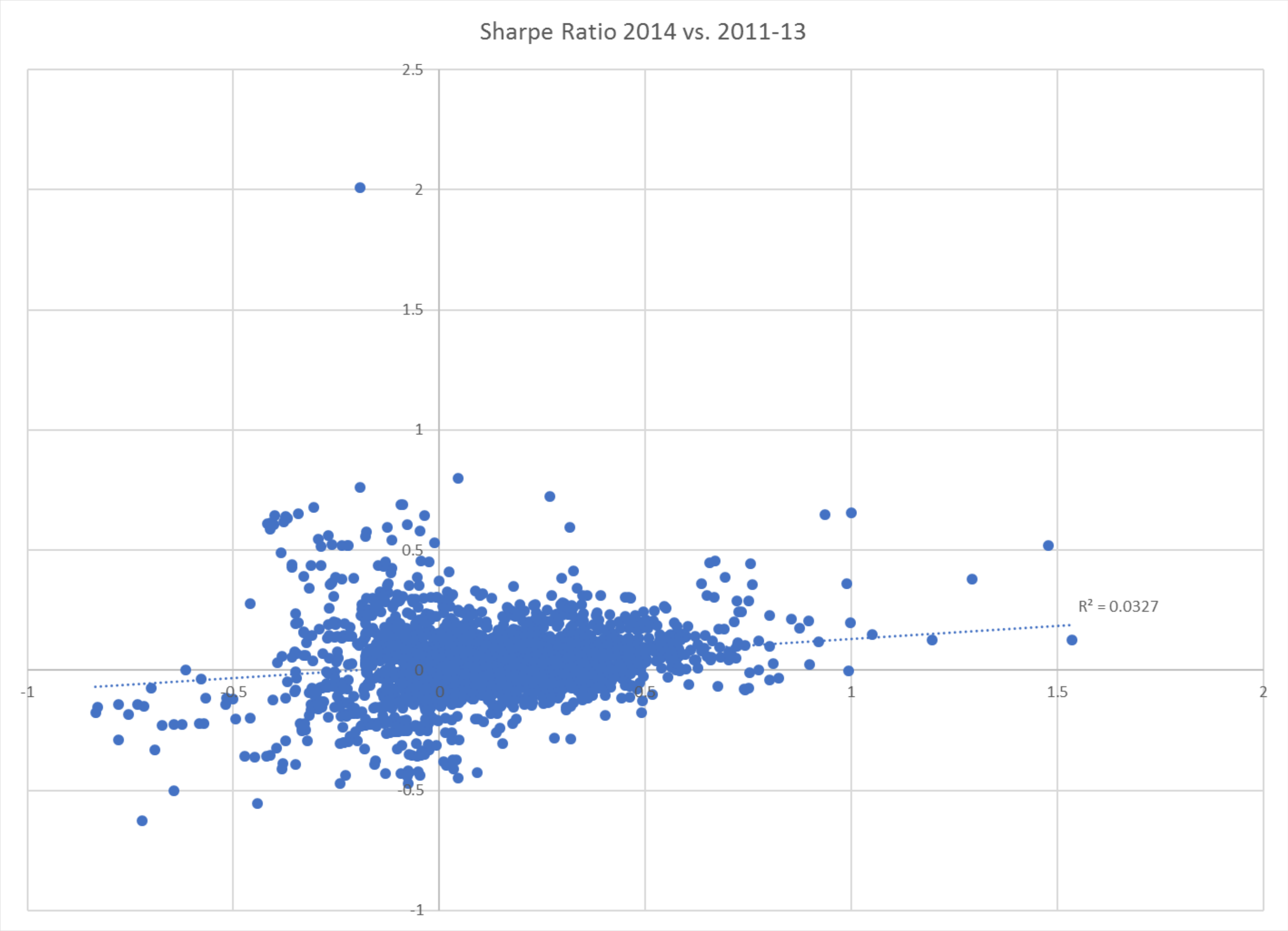
Selection by the Sharpe Ratio?



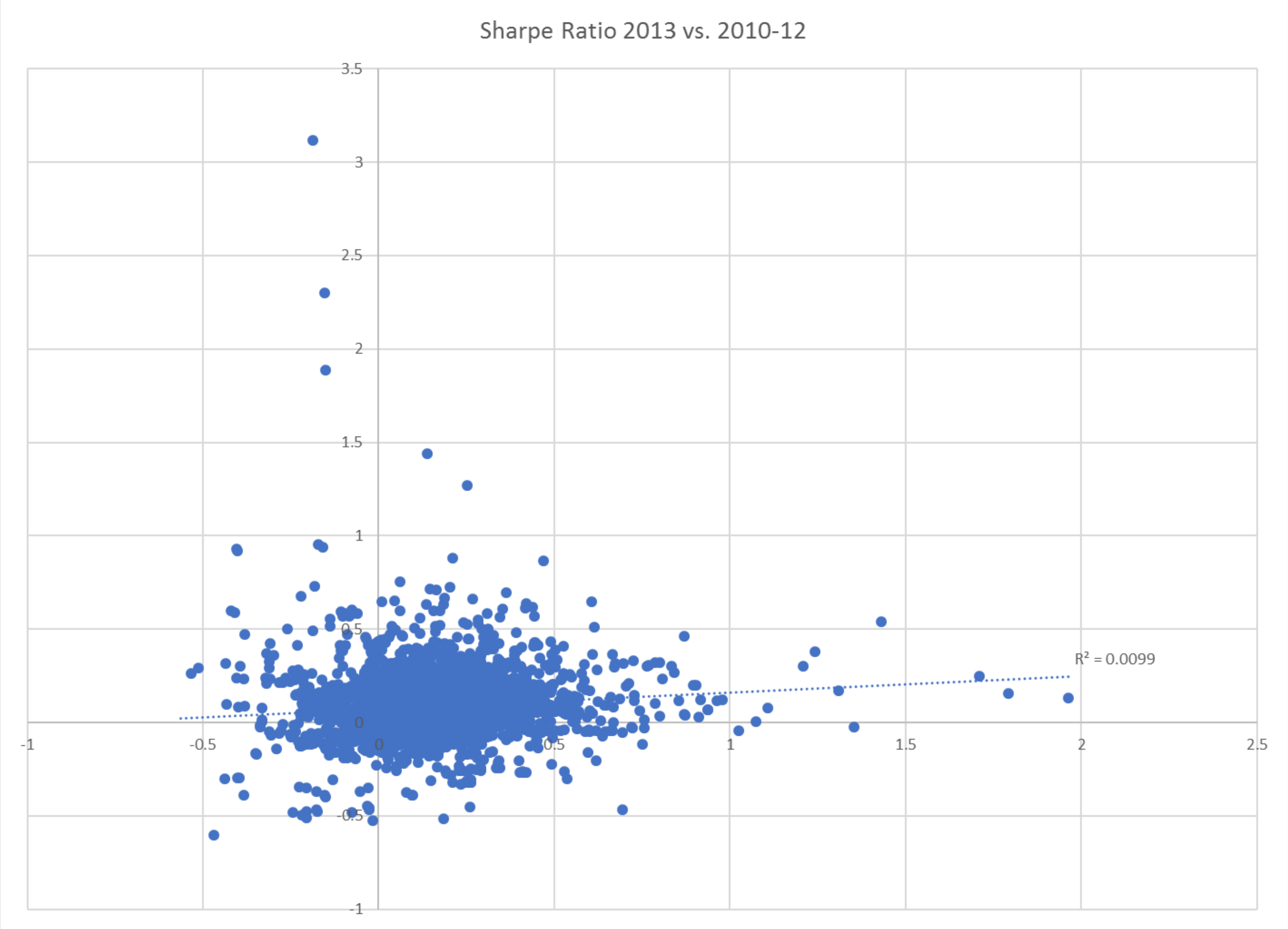
Selection by the Sharpe Ratio?



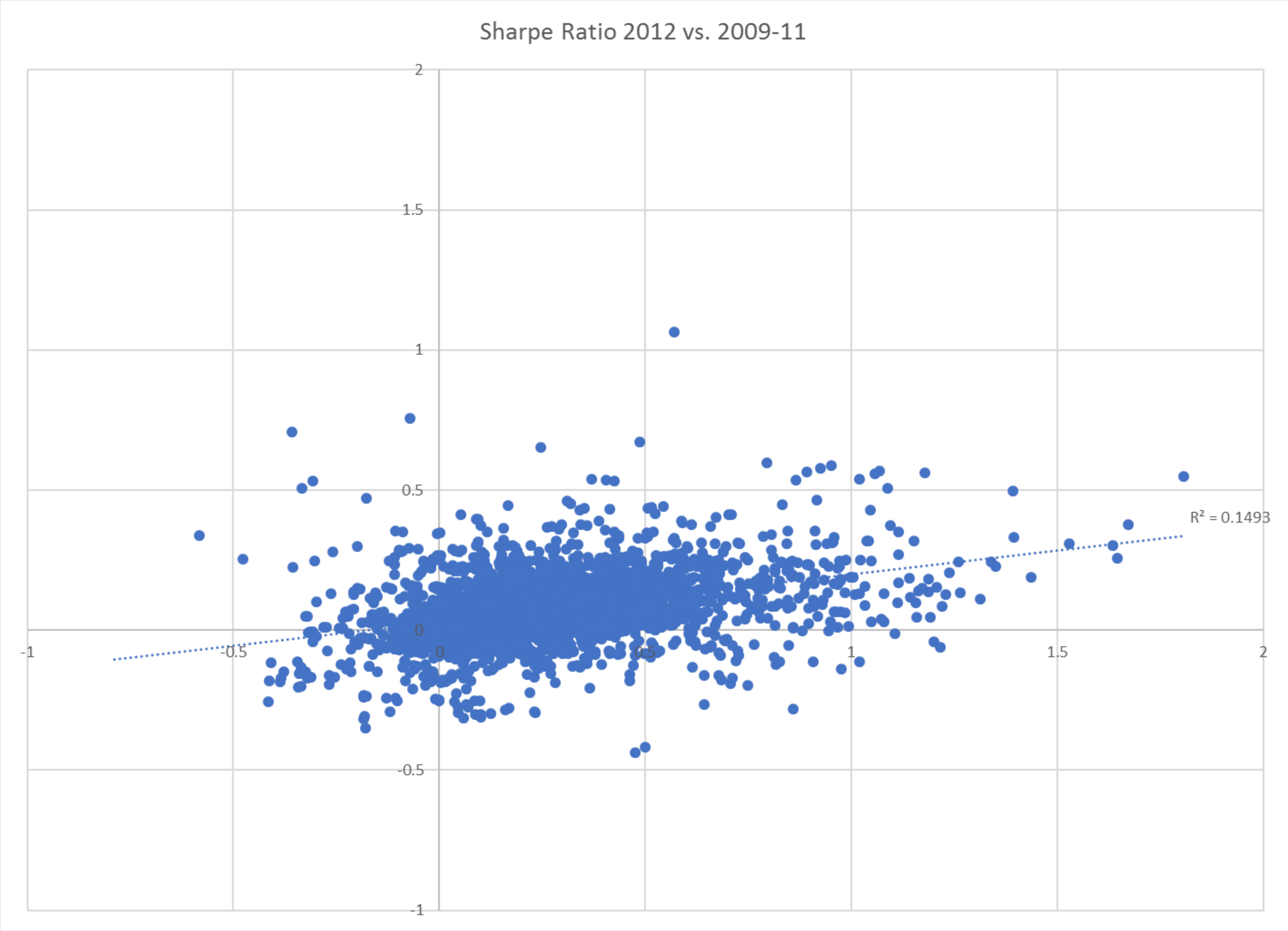
Selection by the Sharpe Ratio?



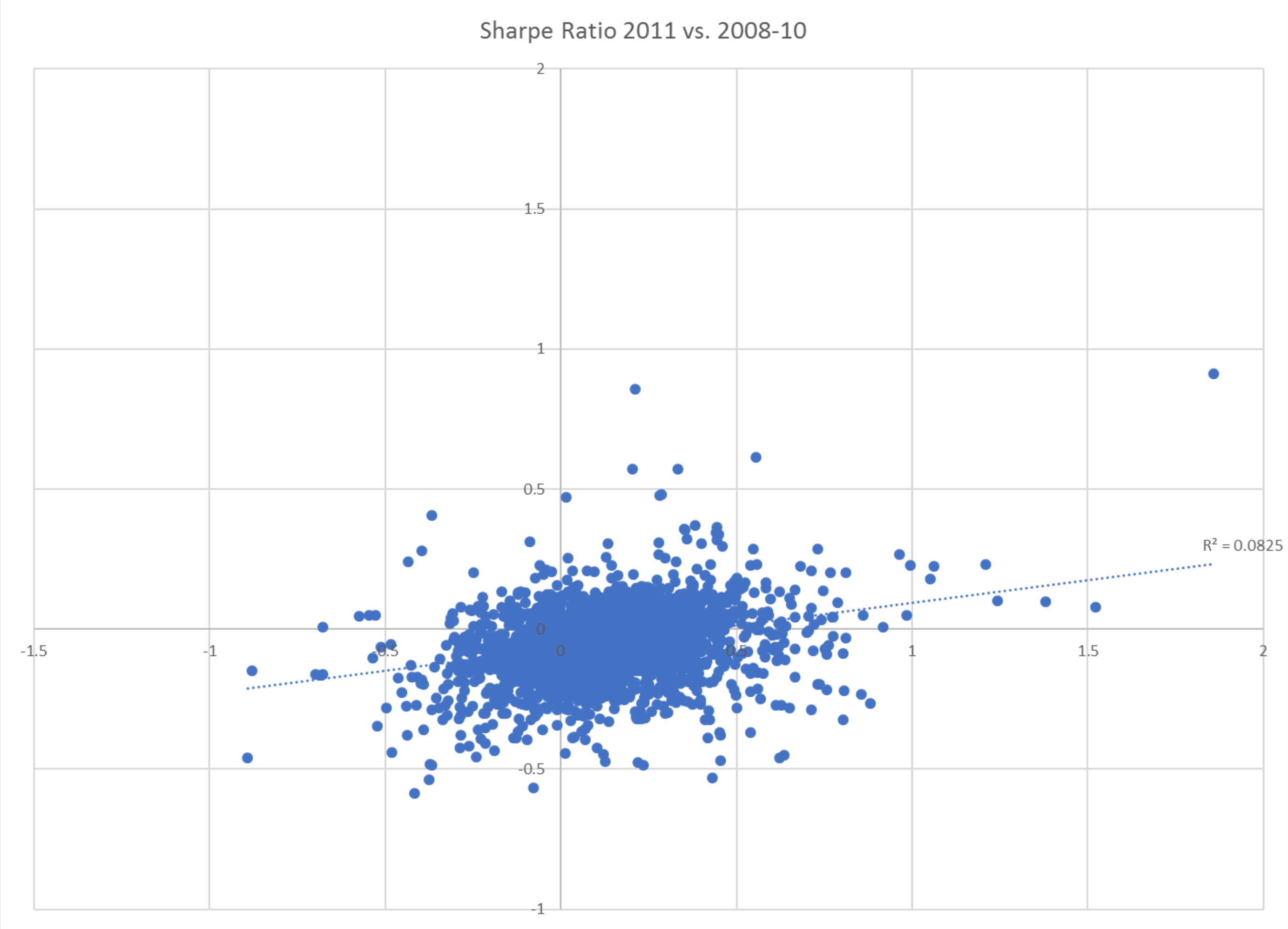
Selection by the Sharpe Ratio?



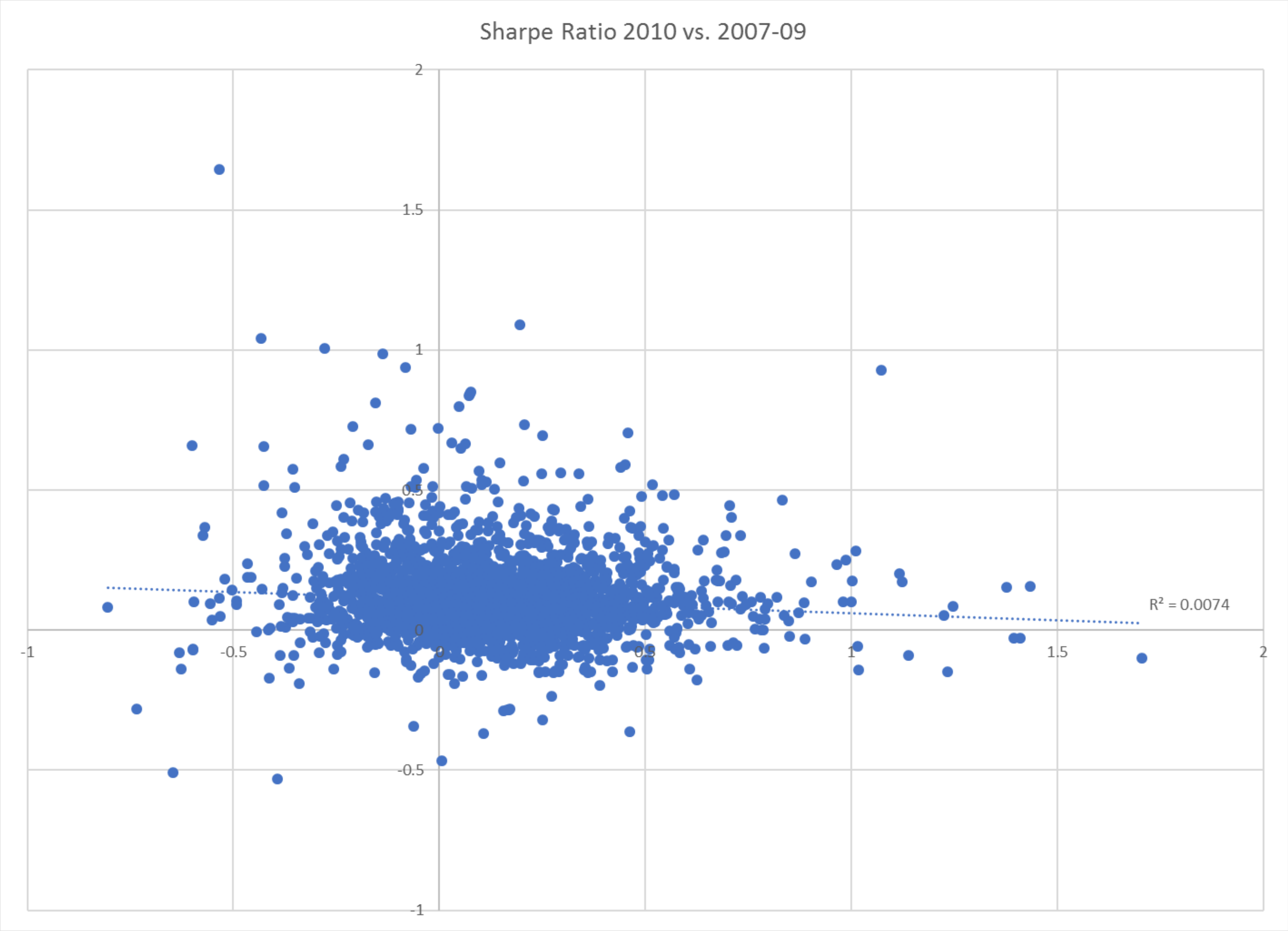
Selection by the Sharpe Ratio?



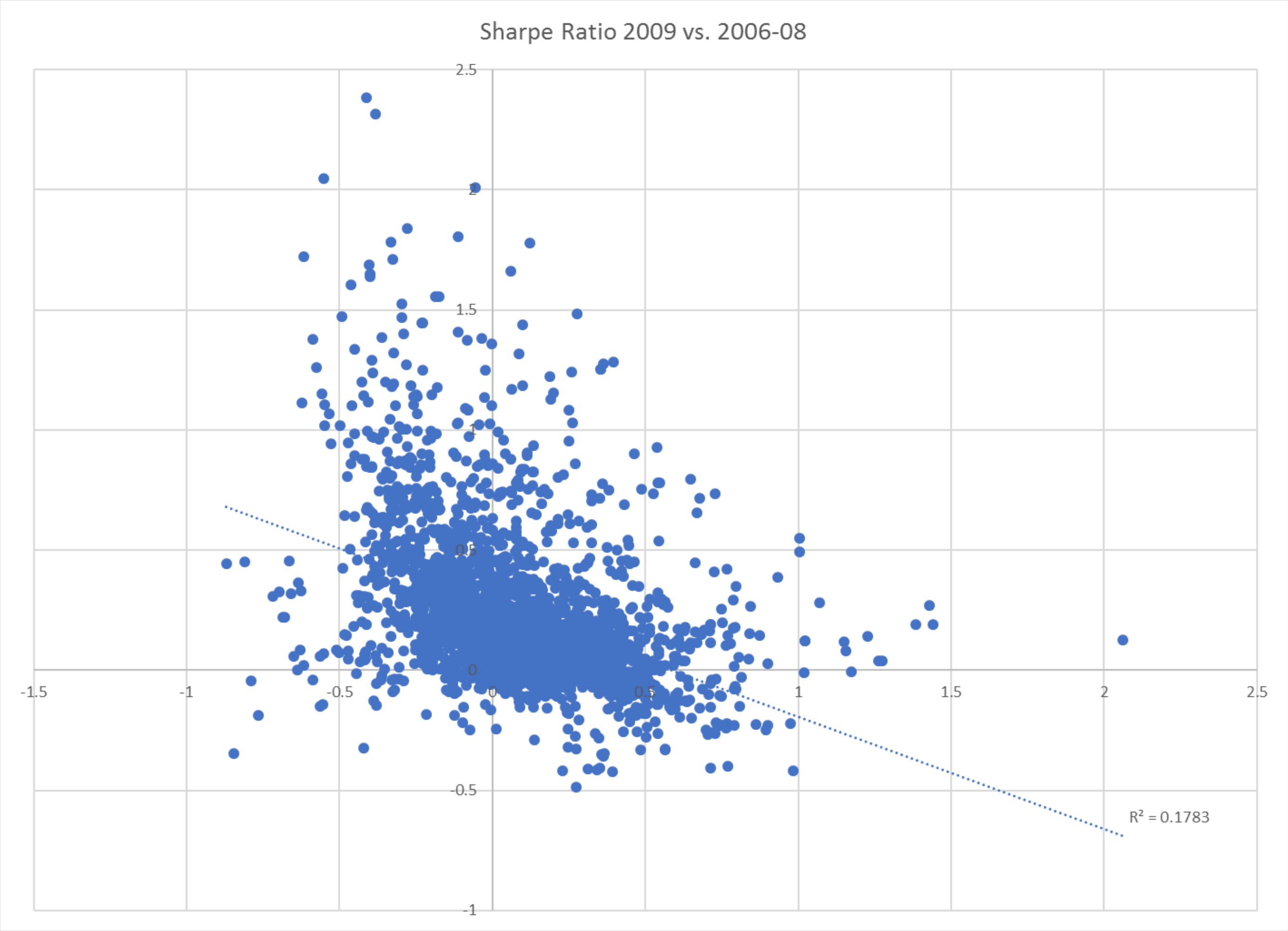
Selection by the Sharpe Ratio?



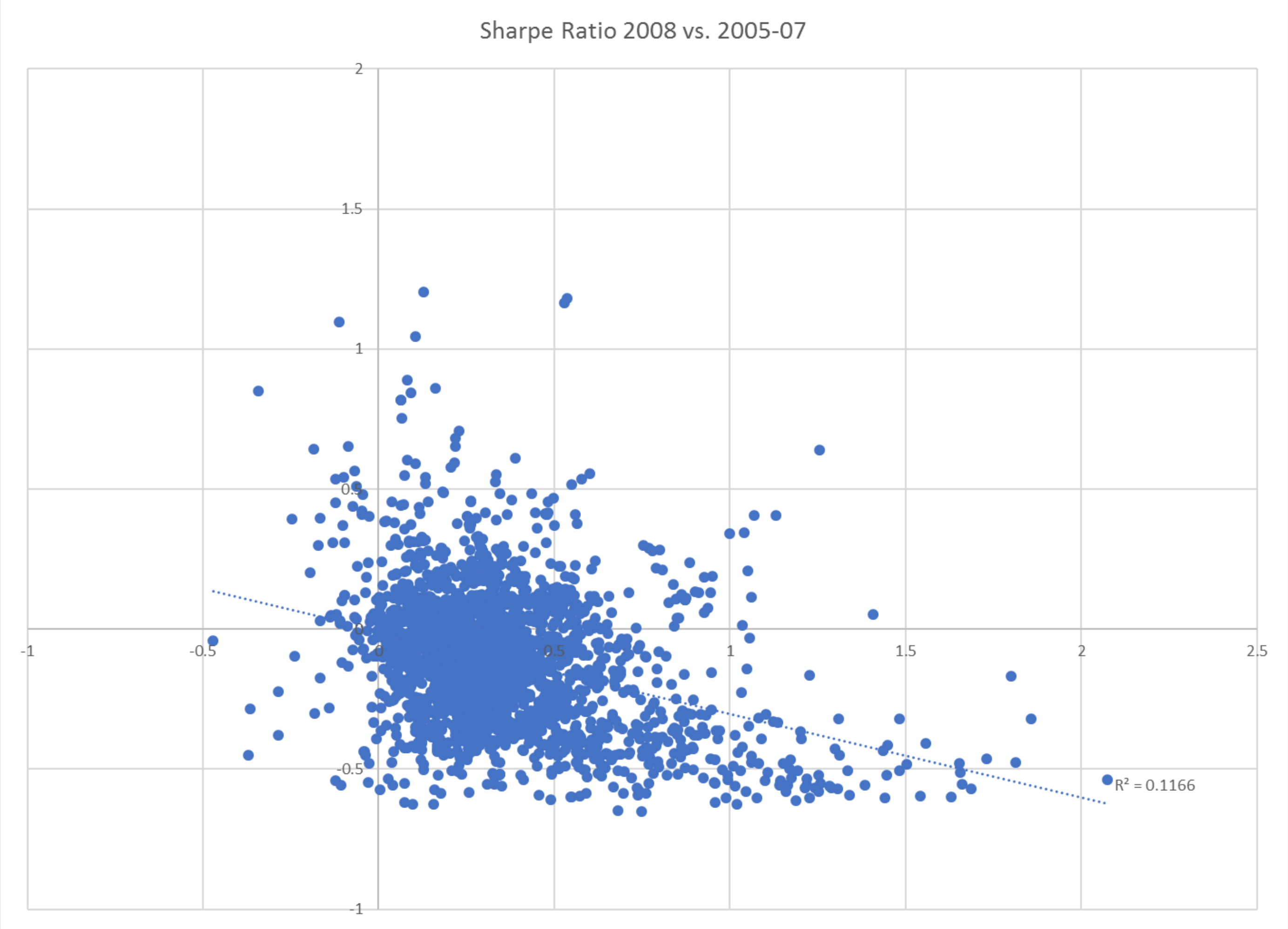
Selection by the Sharpe Ratio?



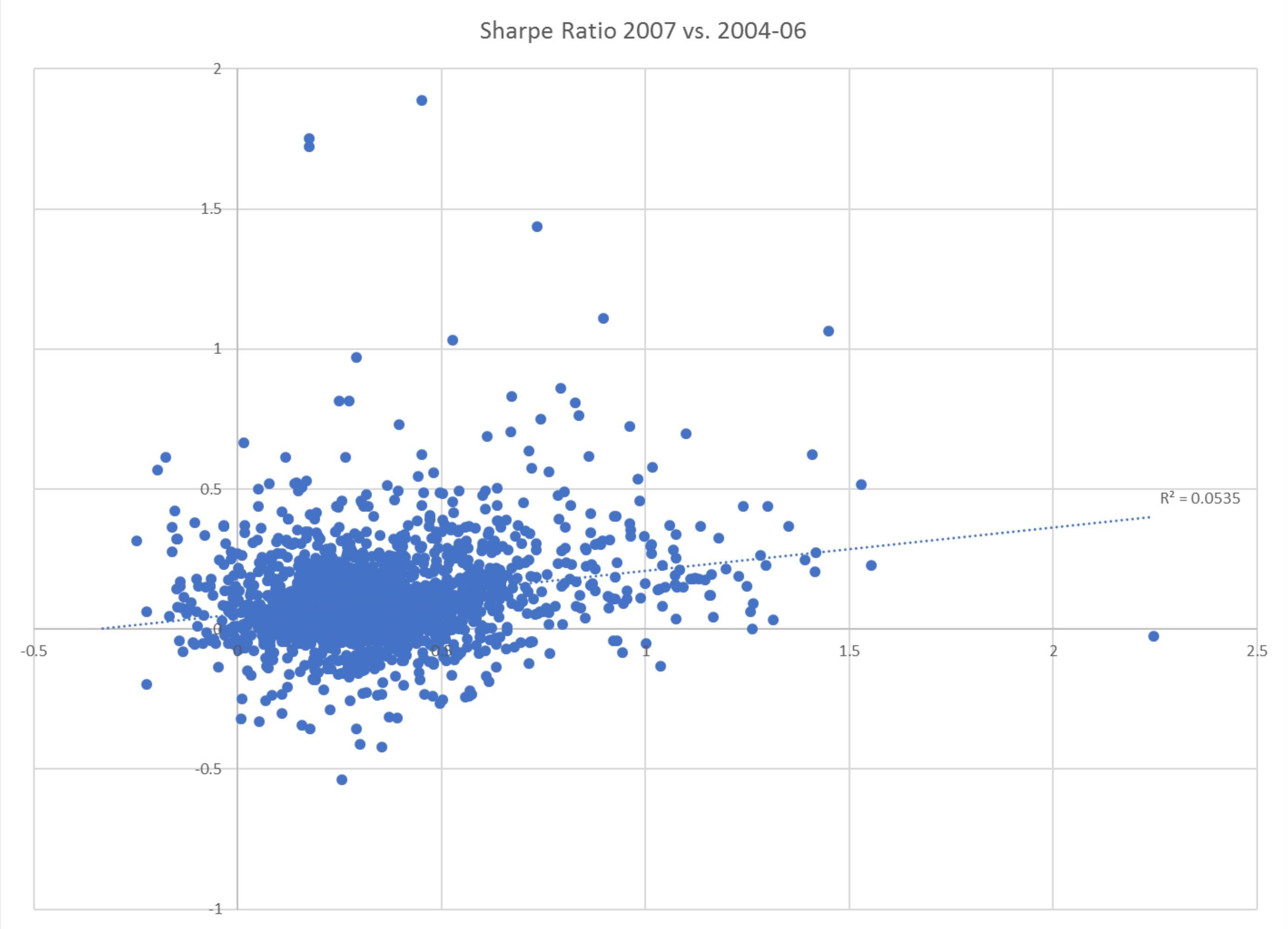
Selection by the Sharpe Ratio?



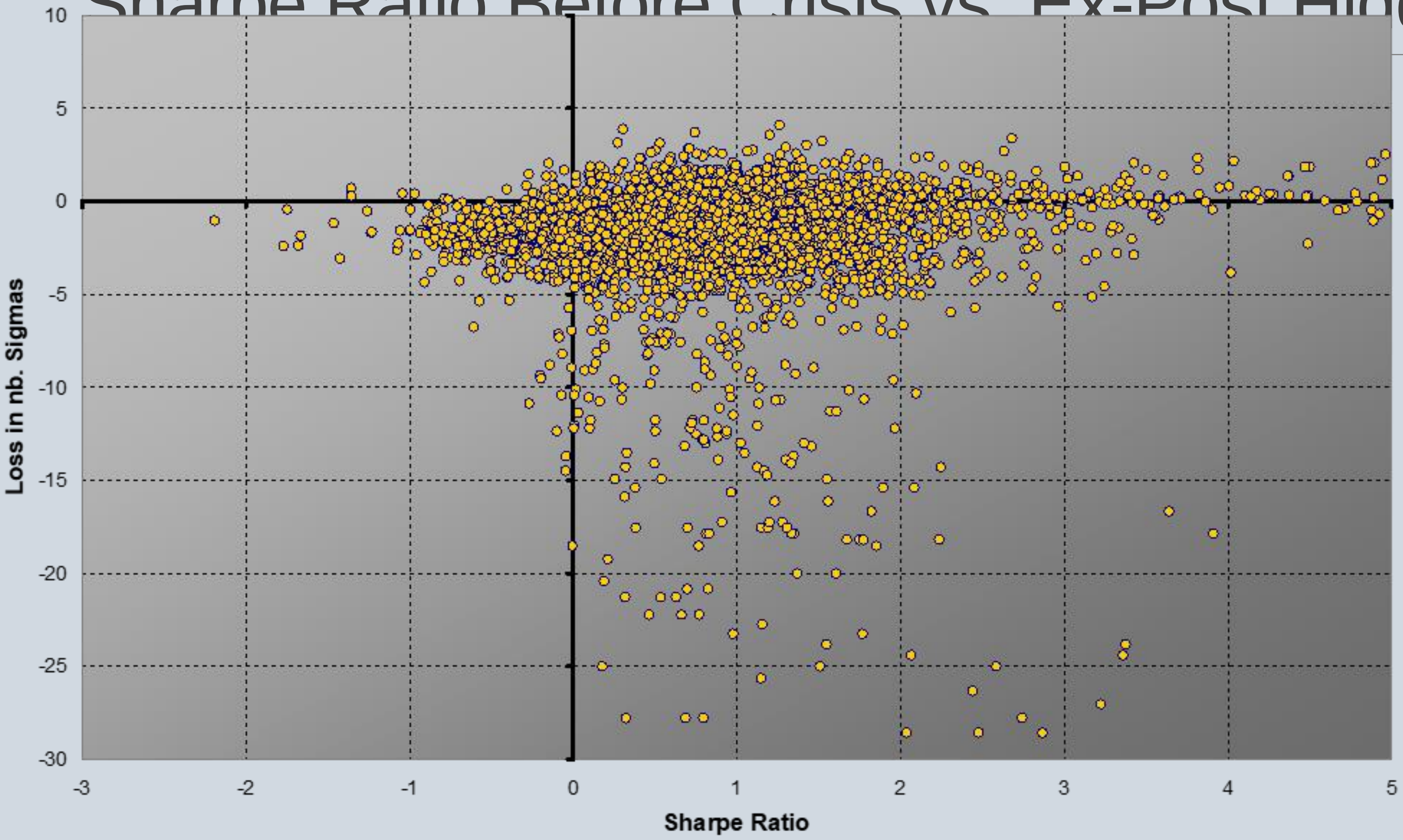
Selection by the Sharpe Ratio?



Selection by the Sharpe Ratio?



Sharpe Ratio Before Crisis vs. Ex-Post Hidden Risk



Using a sample of 3,098 funds

- the X axis is the Sharpe Ratio over the period Jan 04 – Dec 07
- the Y axis is the performance during Sep-Oct 08 divided by the volatility prior to the crisis.

Clearly, the Sharpe ratio is a very poor predictor of losses during the crisis!

Why Traditional “Return-Based” Methods Miss Hidden Risks

Source of Hidden Risk	Example	Effect on Sharpe Ratio
Return Smoothing **Fraud**	Illiquid securities	+++ High Sharpe Ratio
Time Bomb **Short Gamma**	Event Driven Sub-Prime	+++ High Sharpe Ratio
Time Bomb **Surf the Trend**	Event Driven Relative Value...	+++ High Sharpe Ratio

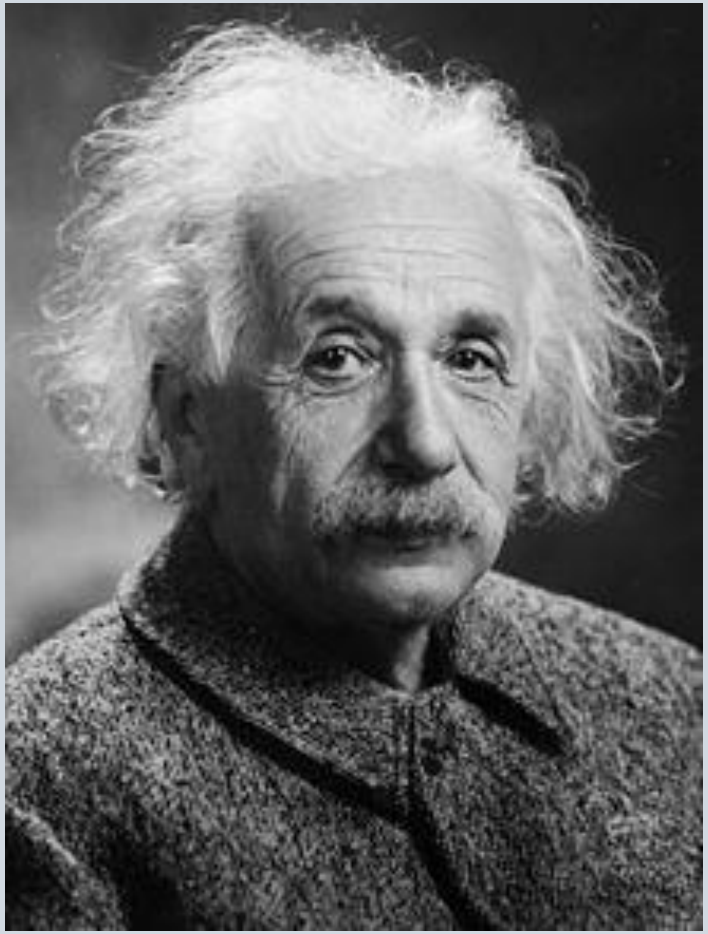
Practically all sources of hidden risks have the effect of boosting the Sharpe ratio.

This explains why past performance is not indicative of future results!

“Time Bombs” refer to typical characteristics of certain trading strategies – those producing small profits a vast majority of the time, but whose occasional extreme losses cancel out years of profits.

For example: Funds that are “short gamma” resemble a strategy that consists of selling a put option on an index and then rolling this position (over years).

Simplicity?



Everything should be as simple as it can be, but not simpler

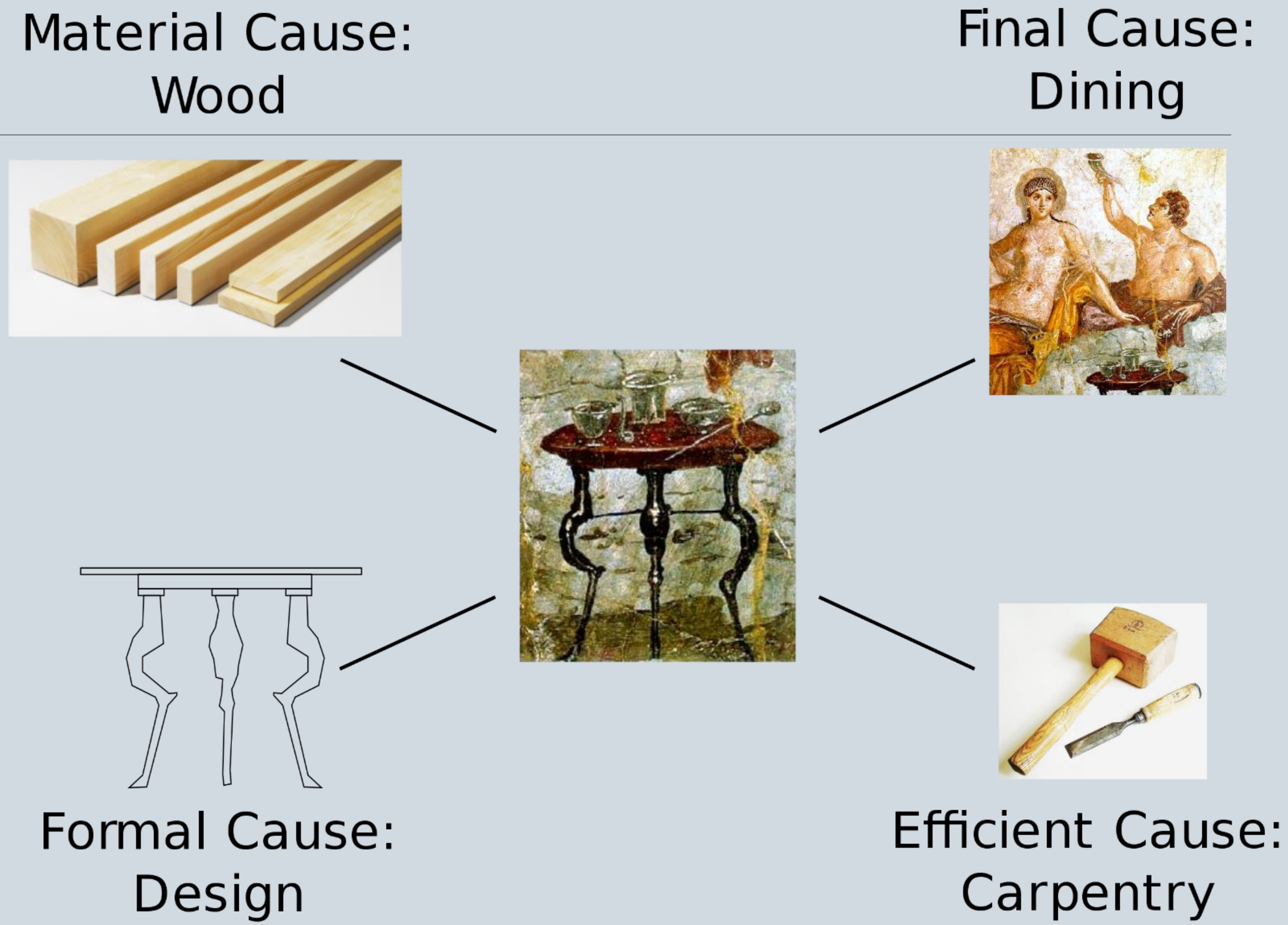
What is « too simple » ?

- People selection: Racism…
- Fund selection: Sharpe Ratio…

« Too simple » is often counterproductive, even as a proxy or a pre-filter

A bit of Phenomenology…

Aristotle’s « 4 causes »



Source: Wikipedia

A bit of Phenomenology

J.F. Courtine: “Phenomenology does not characterize the *"Was"* (what it is), but the *"Wie"* of objects, the how of research, the modality of their "being-given", the way they come to meet.”

We *deal* with managers, we are not zoologist trying to describe each animal, but *actors* in the jungle.

We are exposed to the environment and must cope with it, both with its *opportunities* (we are *predators*) and its *dangers* (we are also *preys*)…

When analyzing funds and managers, knowing the content and the strategy is only a *mean* to understand their *behavior*.

Other approaches to information on the behavior: Statistics and Factor analysis

Math are *very rich*. They are wrong only when their scope is artificially reduced and misses the main point!

Traditional Multi-Factor Analysis

Y = Dependent random variable

$X_1,\ldots,X_n$ = Independent variables

Explanation: $Y = f(X_1,\ldots,X_n) + \varepsilon$

Ideally: $f(X_1,\ldots,X_n) = \mathrm{E}[Y|X_1,\ldots,X_n]$

Statistician's questions:

- Distribution of residuals ε
- Quality of the fit, spurious results?
- Convergence of estimators of f
- Relevance of variables, missing independent variables?

Most often, when $n > 1$, one chooses linear $f \Rightarrow$ Poor prediction power

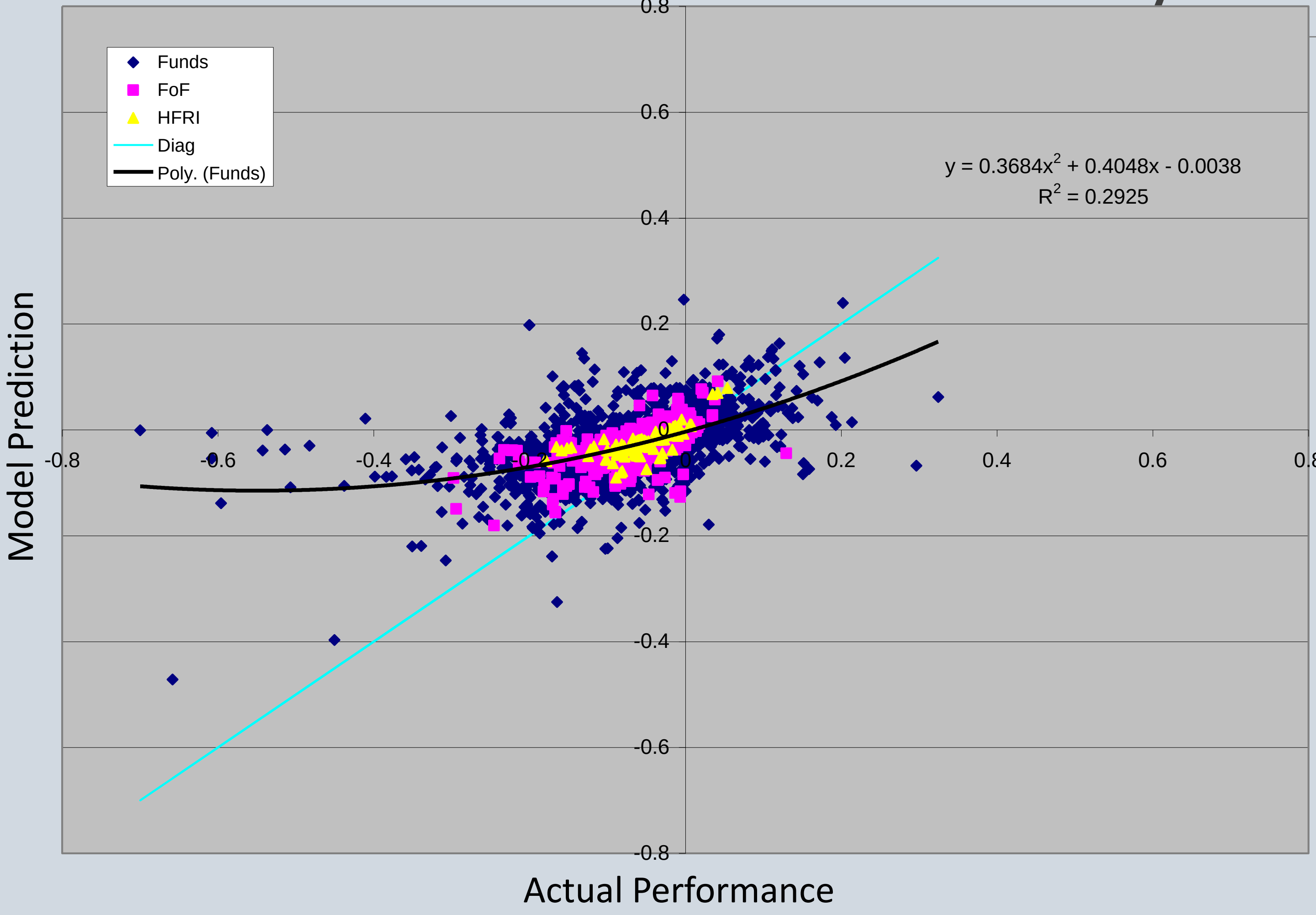
Practitioner's Questions

- What are the sources of risk? Which scenarios are dangerous?
  - Which variables X_i have an influence on Y <u>when it moves dangerously</u>?
  - How far can they go?
- How far can it go? How frequently? How much should I pay (or charge) for an insurance?
  - What is the shape of the return distribution?
  - Fat tails?
- Can I anticipate dangerous events? Can I protect myself against?
  - Crisis prediction (Type I and Type II errors)
  - Hedge efficiency (nature of the hedge, insurer's quality, liquidity risk…)
- Diversification
  - Portfolio return distribution
  - Contribution of each component to the overall risk
  - Portfolio optimization

Problems with Multi-Factor Analysis

- Regime changes invalidate calibration
  - Model f_t calibrated on $[t - \tau, t]$
  - Conditional prediction $\hat{Y}_{t+1} = f_t(X_{1,t+1},\ldots,X_{n,t+1})$
  - Out-of-sample prediction power: $P2 = 1 - \mathrm{Var}(Y_{t+1} - \hat{Y}_{t+1}) / \mathrm{Var}(Y_{t+1})$ <u>often negative!</u>
  - Reasons: model rigidity (linearity), spurious calibration (overfitting), fitting instability (collinearity)
- "Useful most of the time… except when needed!"
  - Question is <u>when</u> is the model being used $\Rightarrow$ Calibration conditionally to usage (e.g. LOWESS)
  - Problem: usage conditions are rare events
- Vanishing Diversification
  - Correlations depend on the regime
  - Optimized portfolios have fatter tails than non-optimized ones

Multi-Factor Error Analysis

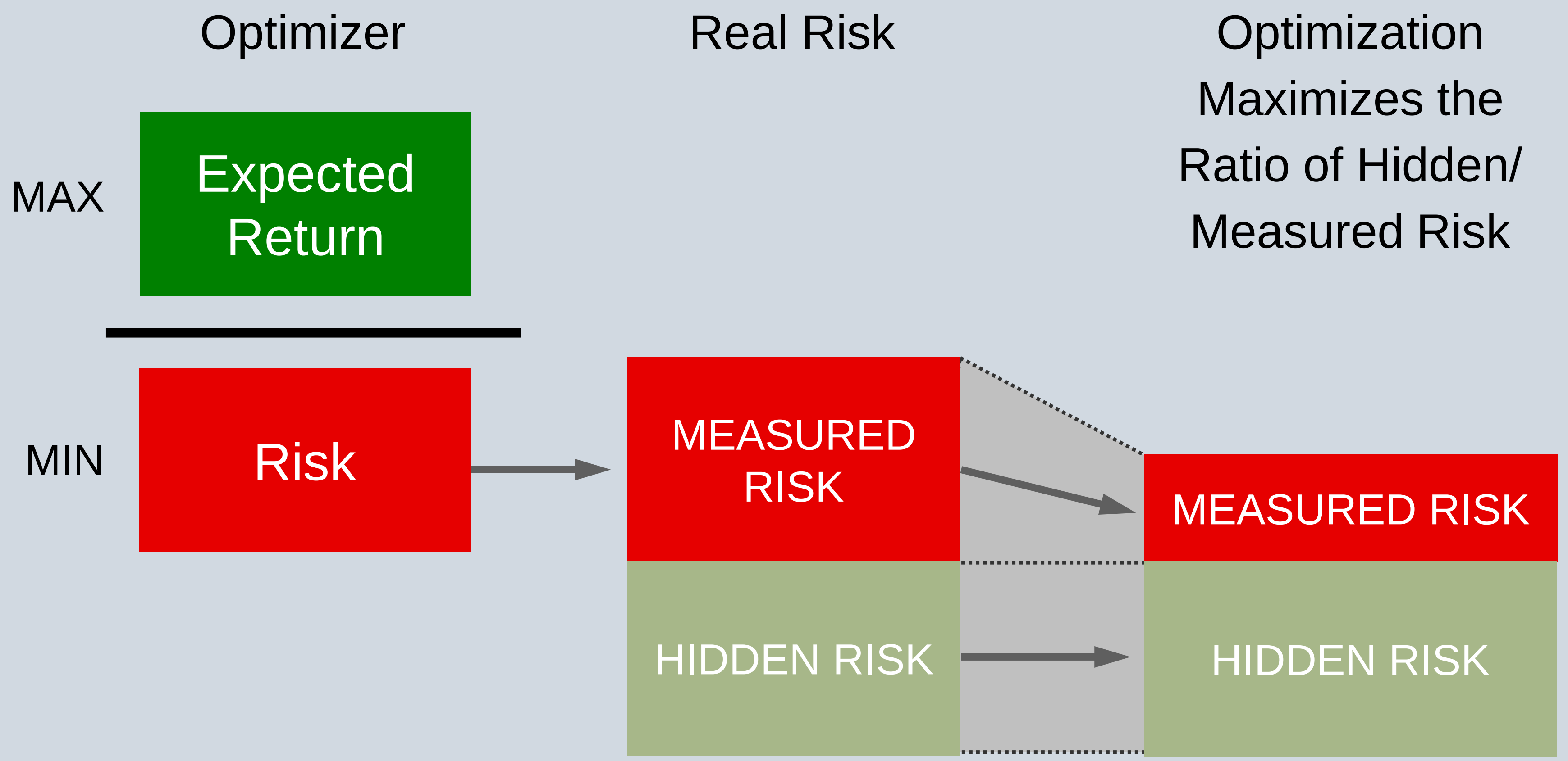


Fung-Hsieh 7 factors model:
Misses large negative events

3000+ fund returns in Sep 08
Coefficients calibrated on [Sep 05, Aug 08]

Data from HFR

Optimizers Failed, However Advanced...



Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Vanishing Diversification

Optimizers, however sophisticated, simply maximize expected return while minimizing *measured* risk. Therefore, by design, optimizers maximize the proportion of unmeasurable risk – i.e. hidden risk – leading automatically to portfolios which eventually deliver very nasty surprises….

Assume that $Y_1 \ldots Y_m$ have mixed joint distribution $P = \pi_1 P_1 + \ldots + \pi_q P_q$ with $\pi_1 > \ldots > \pi_q$

Fat tails can be measured as the ratio of risk under P_1 vs. risk under other regimes.

An optimizer that only accounts for *some* regimes will reduce the risk under those regimes, but *increase* the risk under other regimes, hence increasing fat-tailedness

$\Rightarrow$ Regime changes have a aggravated impact on portfolio risk

Hidden Markov Model

Dynamics in the Market Variable Space

- Finite Number of Regimes $R_1,\ldots,R_m$
  - Covariance Matrix Γ_k
  - Drift Vector μ_k
  - $R_k \sim N(\mu_k,\Gamma_k)$
- In Discrete Time: Transition Probability Matrix
  - $P(t,t+1) = (p_{hk})$
  - p_{hk} = Probability of transitioning from R_h to R_k
  - $\forall h, \quad \sum_{k=1}^{m} p_{hk} = 1$
  - $P(t,t+n) = P^n$
- In Continuous Time: Semi-group Generator A
  - $P(t,t+s) = e^{sA}$

Simulation: Markov Chain Monte-Carlo

Euler Scheme:

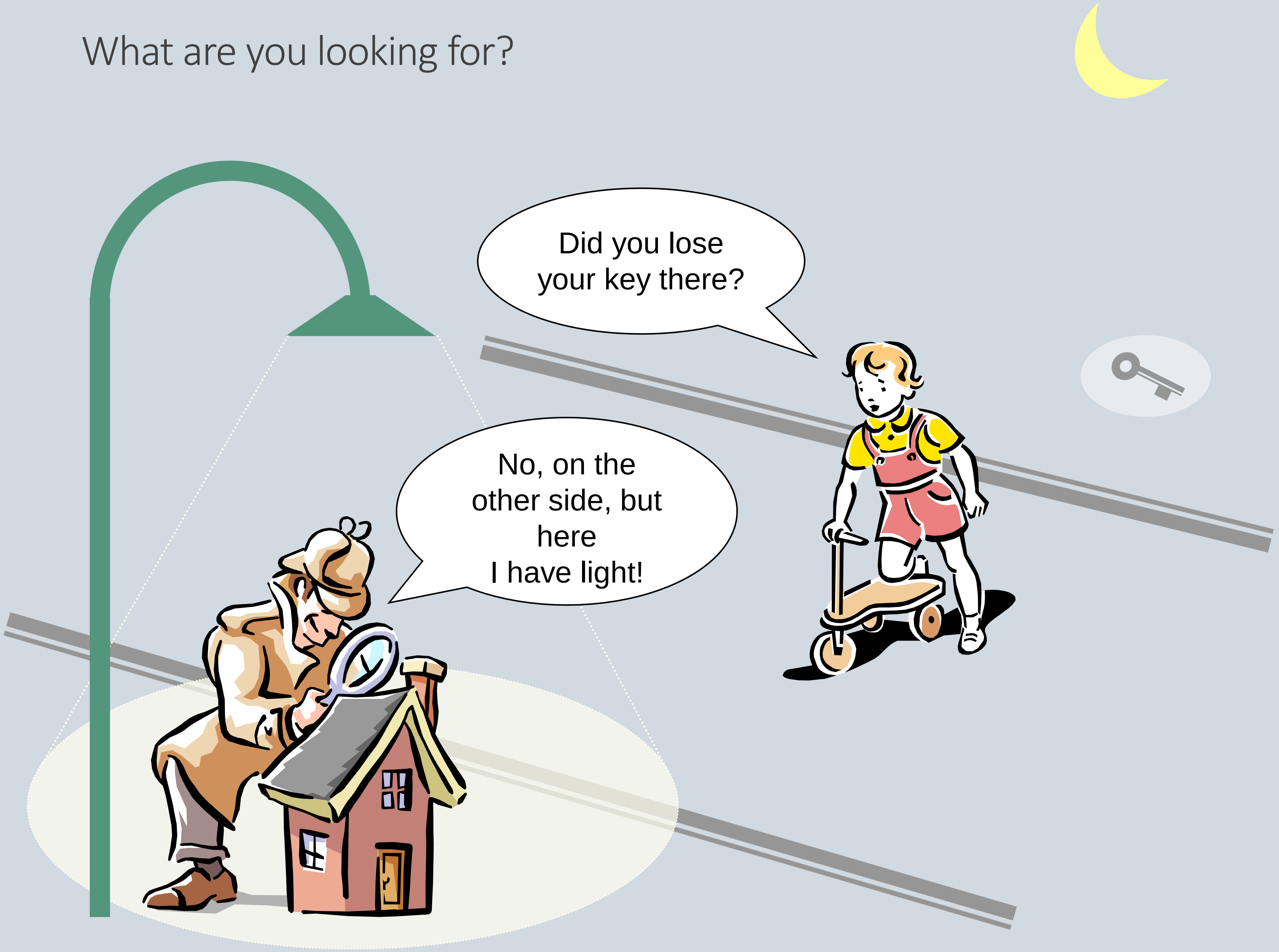
- Discretized Time $t_0,\dots,t_n$
- Pick new regime $R(t_{i+1})$ according to P applied to current regime $R(t_i)$
- Simulate Market Evolution following $R(t_{i+1})$

Gaussian Mixture $\Rightarrow$ “Fat Tails”

Asymptotically Gaussian

- $\mu_\infty = \sum_{k=1}^{m} \pi_k \mu_k$
- $\Gamma_\infty = \sum_{k=1}^{m} \pi_k \Gamma_k$
- $P'\pi = \pi \quad \pi = (\pi_1,\dots,\pi_m)$

What are you looking for?



The Data Wall

- Millions of Assets and Funds
  - A few years of history => only a few 10's to a few 100's of returns
  - Regime changes: future ≠ past
  - Position info: unreliable, incomplete, delayed, fast changing
  - Large variety of strategies and trading universe
- 10,000's Risk Factors
  - All asset classes
  - Long term history, including many crises, cycles
  - Hedge Funds often uncorrelated to markets: need exotic factors
  - Correlations only appear during crises: need nonlinear models

Too many models, too little information

IMPOSSIBLE TO **SELECT** AND **CALIBRATE** A MODEL

Polymodel Principle

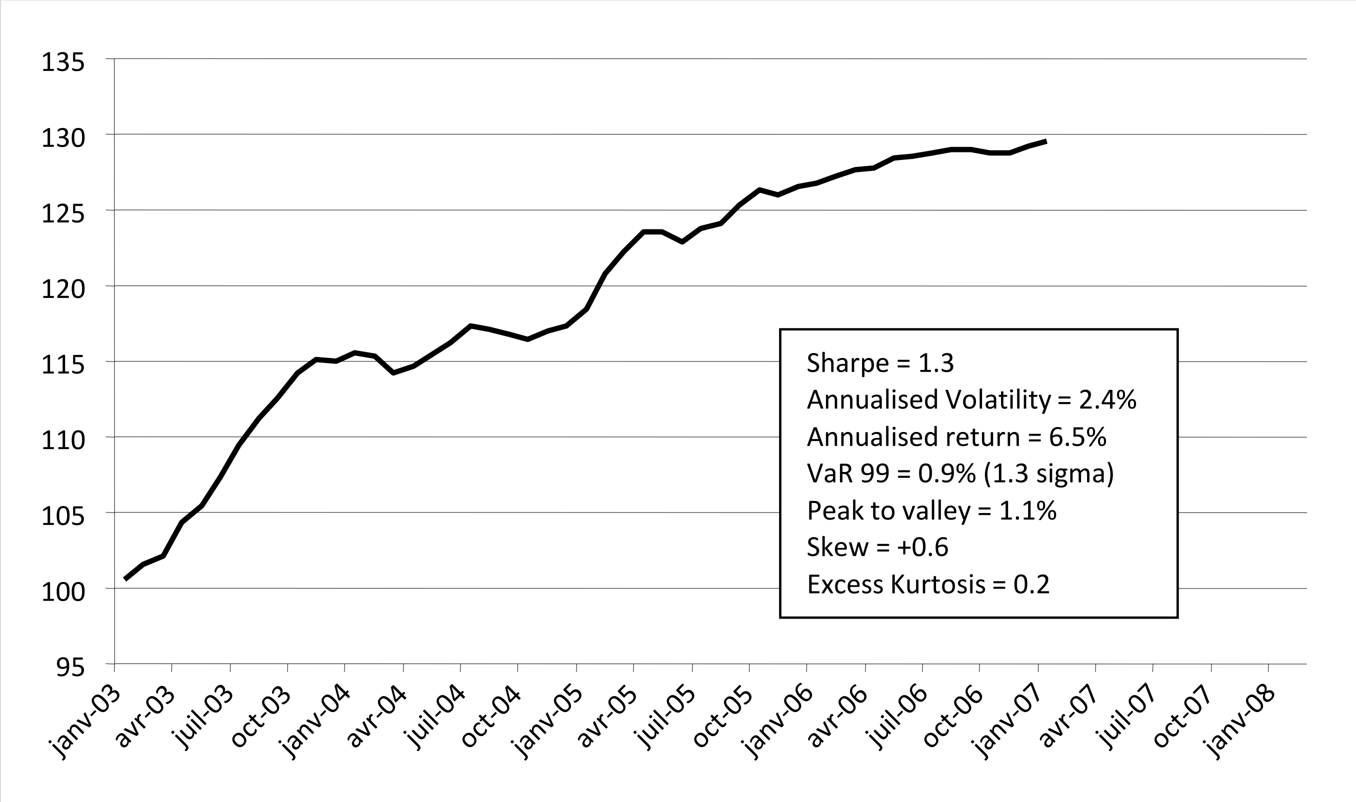
- Replace one function $Y = f(X_1,\ldots,X_n) + \varepsilon$ by a collection:
  - $Y = f_1(X_1) + \varepsilon_1$ $\qquad f_1(X_1) = \mathrm{E}[Y|X_1]$
  - …
  - $Y = f_n(X_n) + \varepsilon_n$ $\qquad f_n(X_n) = \mathrm{E}[Y|X_n]$

  Taken individually, each model is imprecise, but the collection contains at least as much information as a multi-factor model, and in fact much more !

- Individual models can be nonlinear and contain lags
  - Easy to calibrate
  - Focus on where useful info lies: in the extremes

- Factors can be correlated and linearly, or even nonlinearly dependent
  - Use hundreds, or even thousands of factors
  - Need to assess the <u>relevance</u> of each factor $\Rightarrow$ p-value

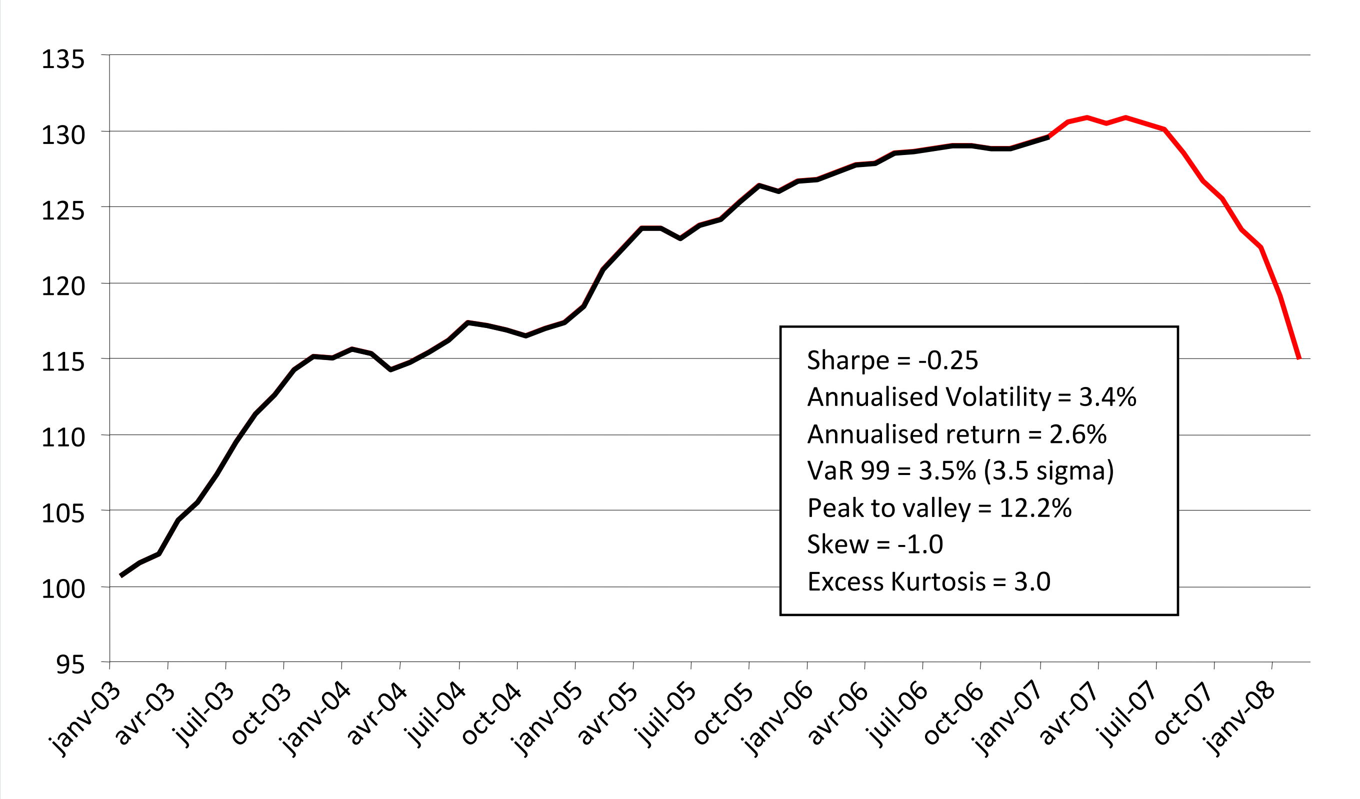
Manager Selection



This performance series would attract any investor who is solely focused on past performances. The sequel shows how the investor might be disappointed.

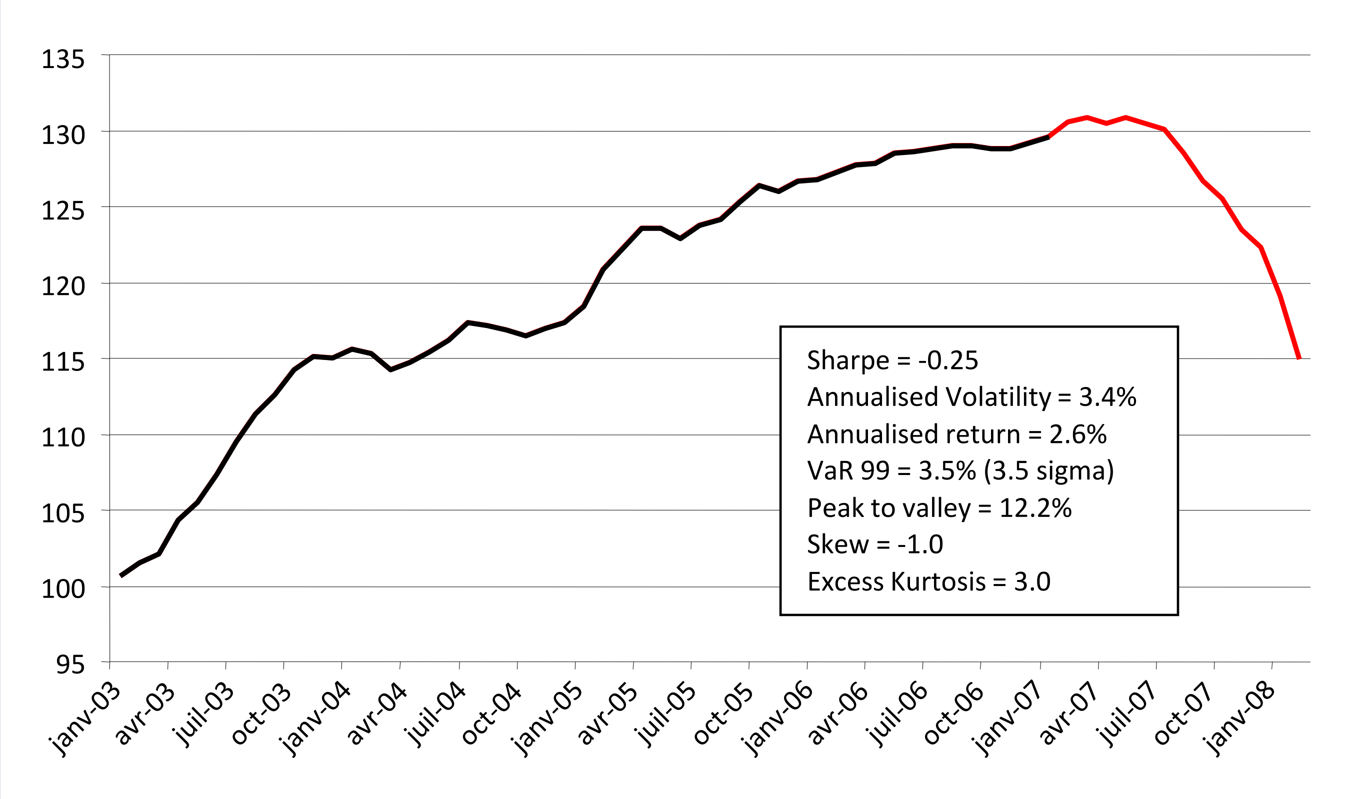
WILL IT LAST?

Manager Selection



Could such a loss be anticipated by looking only at the past fund performances?

Manager Selection



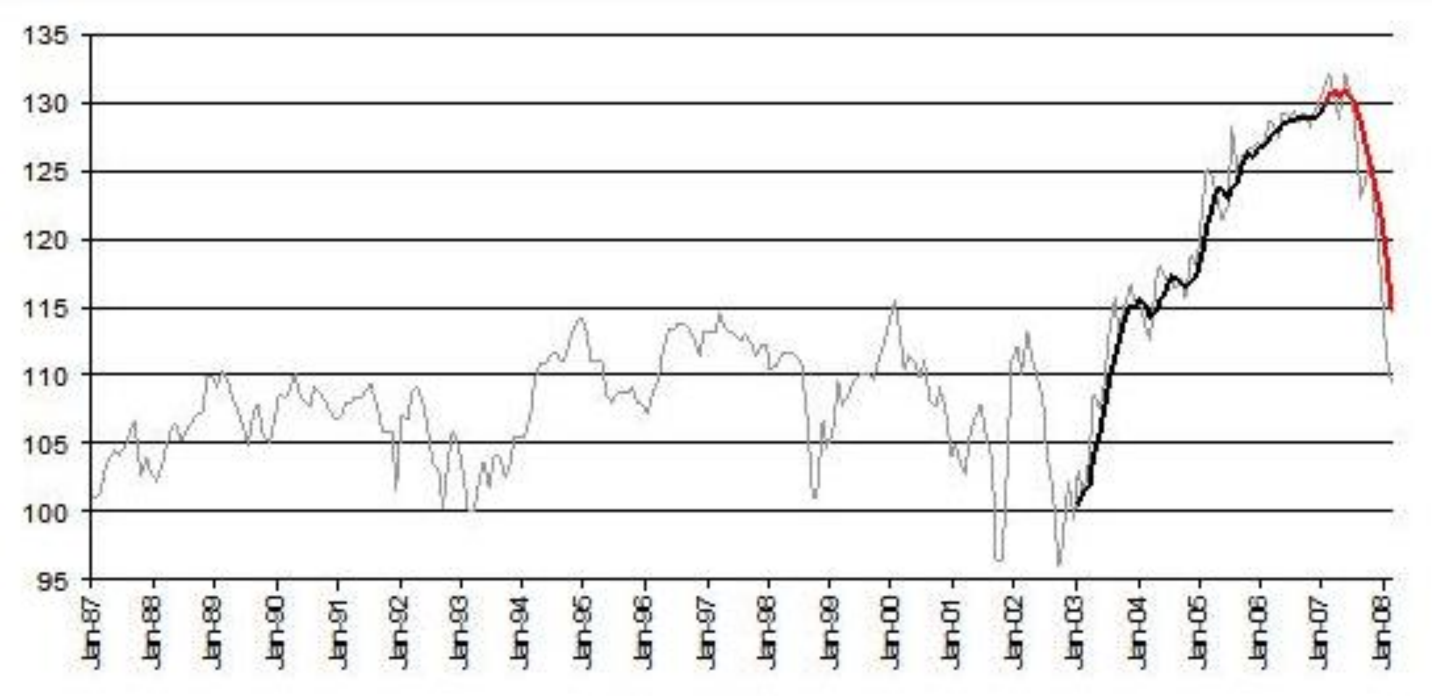
Credit driven fund:

- Long AAA bonds, Short T-bonds, duration 10Y

Could such a loss be anticipated, only looking at past fund performance?

Yes, with factor analysis

Factor Analysis



These fund returns depend mostly on the AAA credit spread – in a nonlinear (optional) manner.

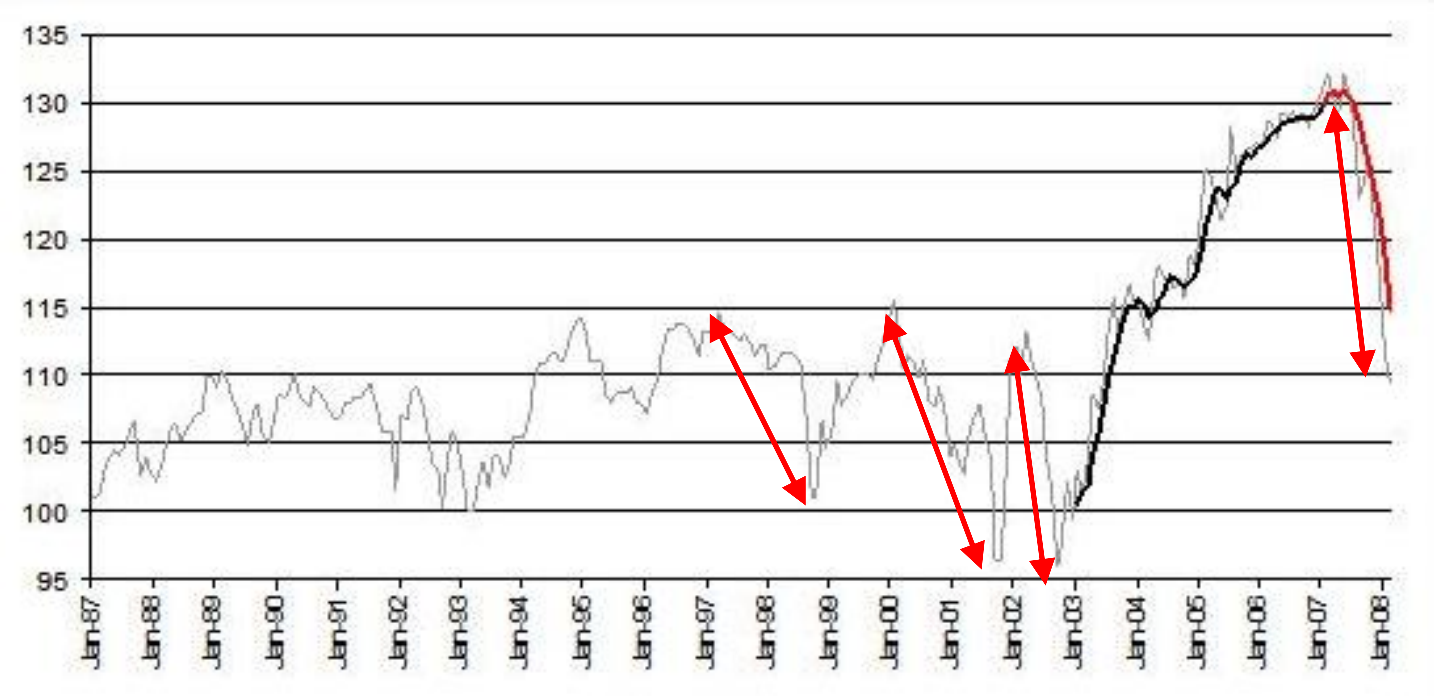
The grey curve is obtained by aggregating the nonlinear function of credit spread changes over many years.

This leads us to a novel approach for anticipating extreme risk, namely through the concept of STRESS VAR.

Credit driven fund vs. AAA spread over T-Bonds:

- This fund was just surfing the good wave during the analysis period

Factor Analysis



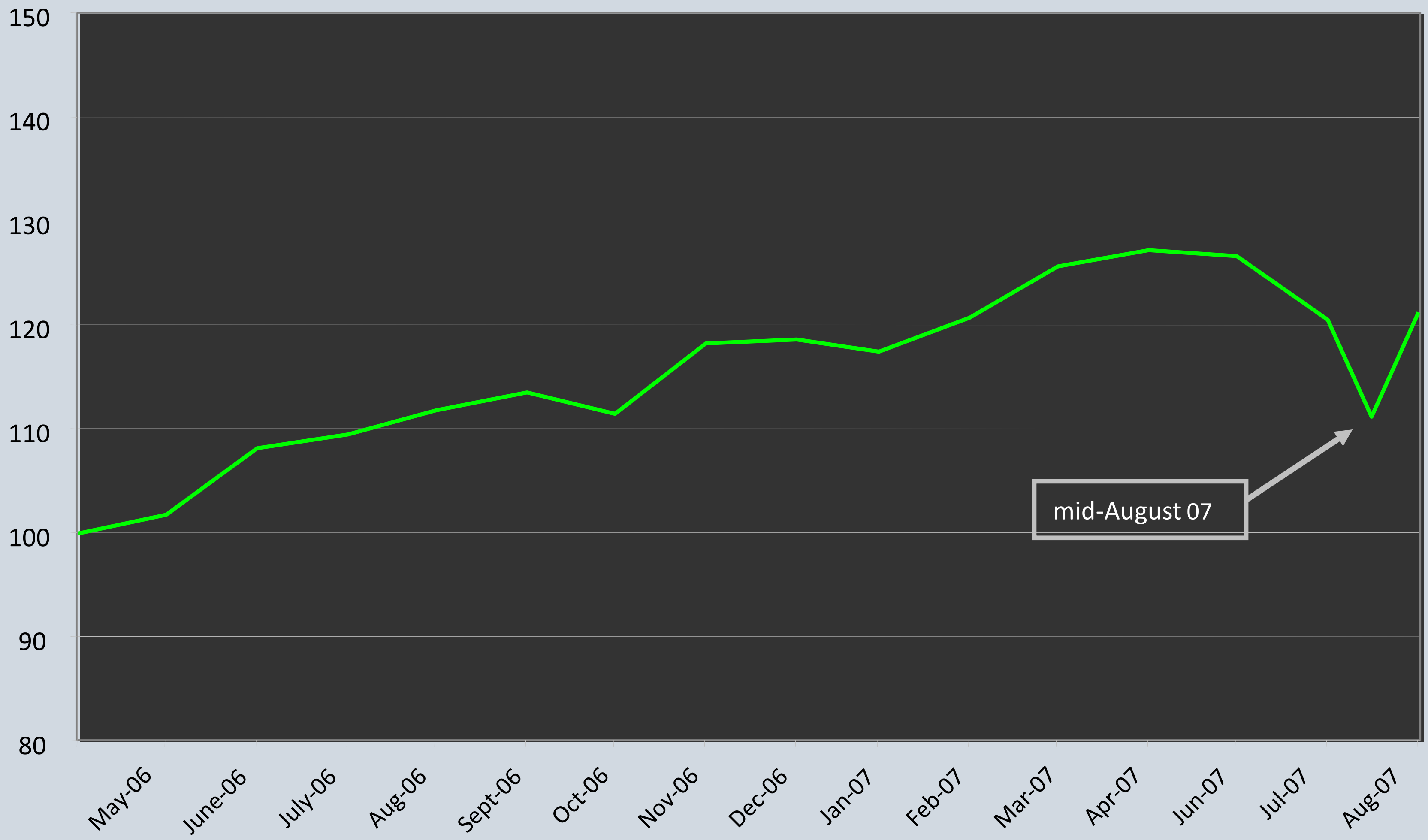
One can see that the loss experienced in 2007 had several similar precedents.

The “Stress VaR” is derived from extrapolated losses of the fund, prior to its actual track record.

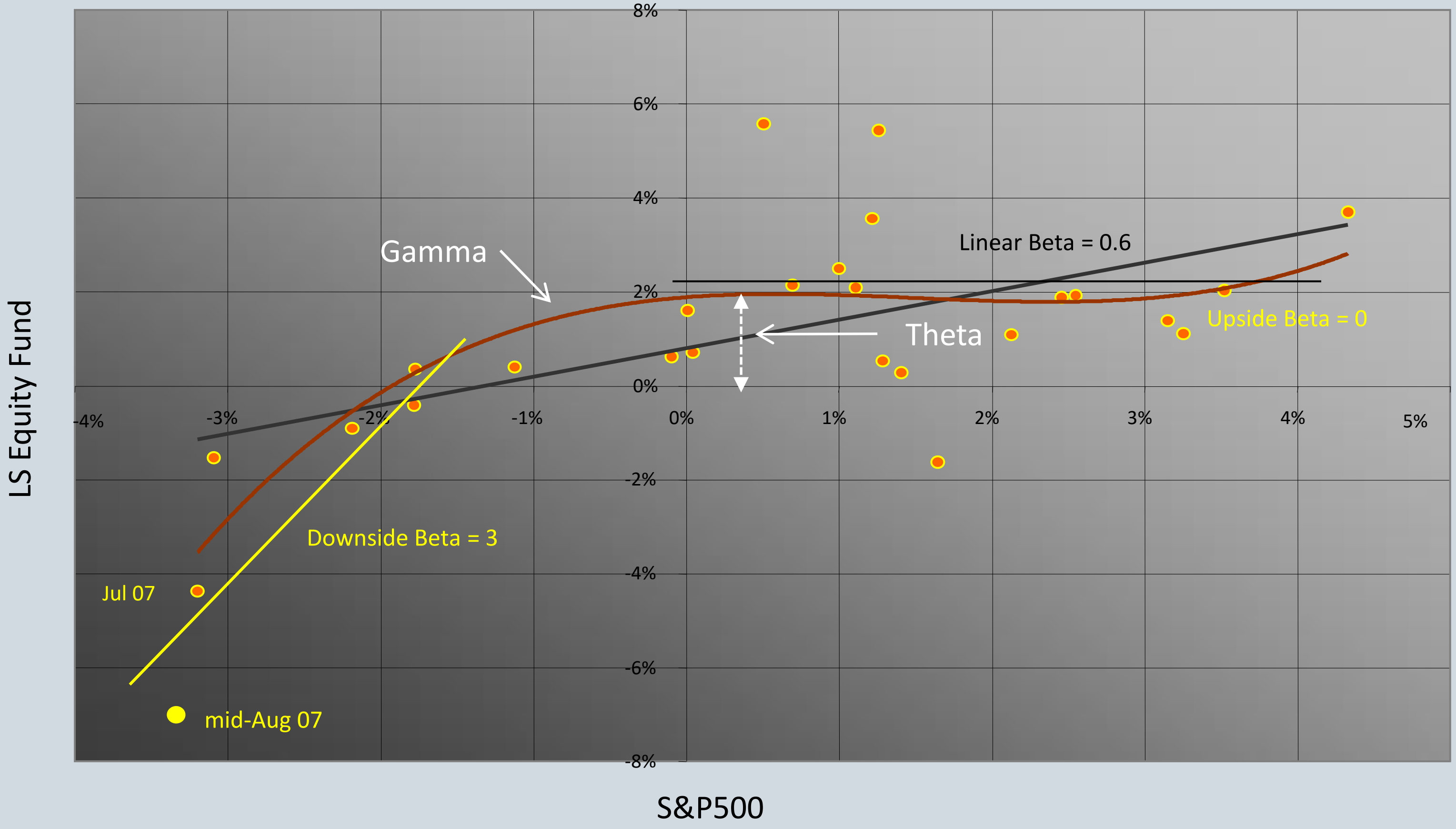
Credit driven fund vs. AAA spread over T-Bonds:

- The driving factor experienced in many past jumps comparable to the crisis

Case Study: Long-Short Equity



Scatter Plot of Returns vs. S&P 500

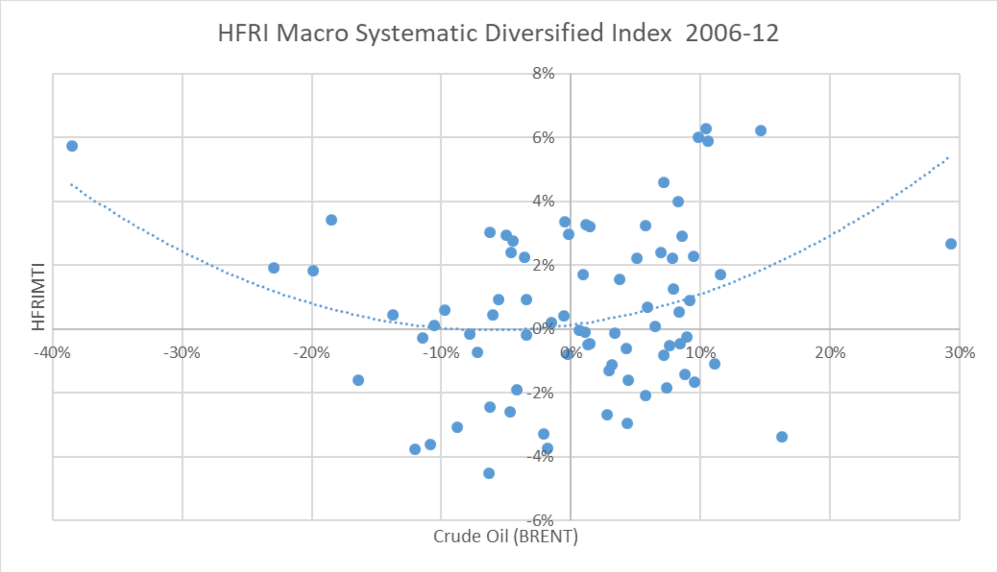


Long-Short Equity Fund Beta change in a *systematic* manner The traditional linear Beta of the fund is 0.6.

But by carefully examining these returns, one finds that the “upside Beta” is 0, while the “downside Beta” reaches 3.

Had the S&P fallen by 15% instead of 5%, the fund would have lost an extra 30%, reaching almost 40%, sufficient to put it out of business.

CTA Systematic: A *Convex* Strategy vs. Oil



Dominant Factor™ Analysis by Nonlinear Polymodels

By definition, a "model" is aimed at *stressing* a fund and anticipate its response to the stress

A *Polymodel* is a Collection of Single-Factor (Nonlinear) Models

Step 1: Identify a **LARGE** Set of Factors

- **LONG HISTORY** (20+ Yrs incl. crises)
- As many factors as potential **risk sources** $\Rightarrow$ **Several 100's**

Step 2: Scan Factors One at a Time

- Focus on **EXTREME MOVES** $\Rightarrow$ **Nonlinear Models**
- Select only factors with a strong statistical relationship to the fund $\Rightarrow$ **Score = –Ln(*p*-value)**
- A risk factor is *dominant* if its *Score* is at least ½ highest score across all factors

Step 3: Stress Selected Factors

- Information Ratio = Impact of Factor / Uncertainty
- For each stress test, select the Factor(s) with the highest information ratio (may depend on the stress test)
- **Merge** single-factor models to maximize Information Ratio

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor™ Analysis by Nonlinear Polymodels

Dominant Factor Analysis is aimed at breaking the "data wall".

- Use short historical data to identify risk sources and fund sensitivity to them
- Use (very) long historical data on risk factors for a thorough assessment of what may happen to the fund (good *and* bad)

Here, the major innovation is in the way that the distribution of future returns is estimated; using a very long history of markets in order to include past crises, a large number of factors in order to account for all possible risk sources and a collection of nonlinear models in order to account for extreme risks – in particular, the impact of *liquidity gaps*. Short fund historical records are utilized in an optimal way.

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Poly-model Structure

Input:
Asset Historical Data Y_i

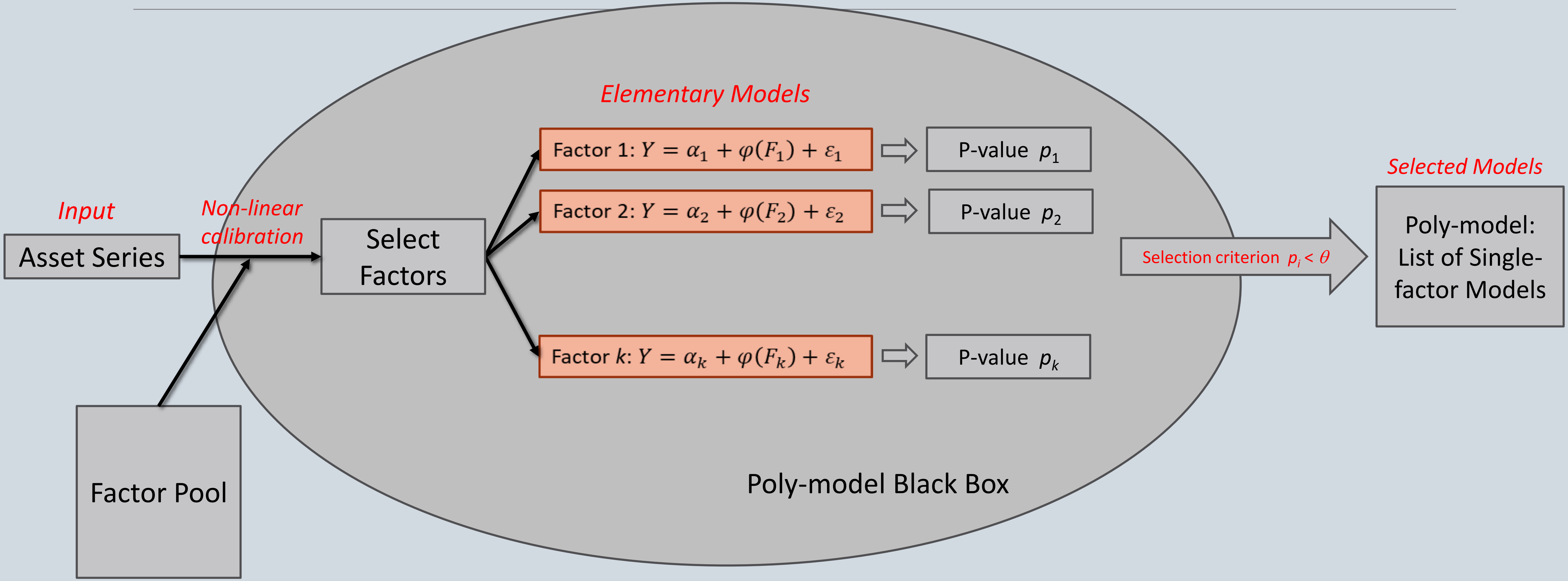
Poly-model Analysis

$Y_i = \varphi_{ij}(F_j) + \varepsilon_{ij}$

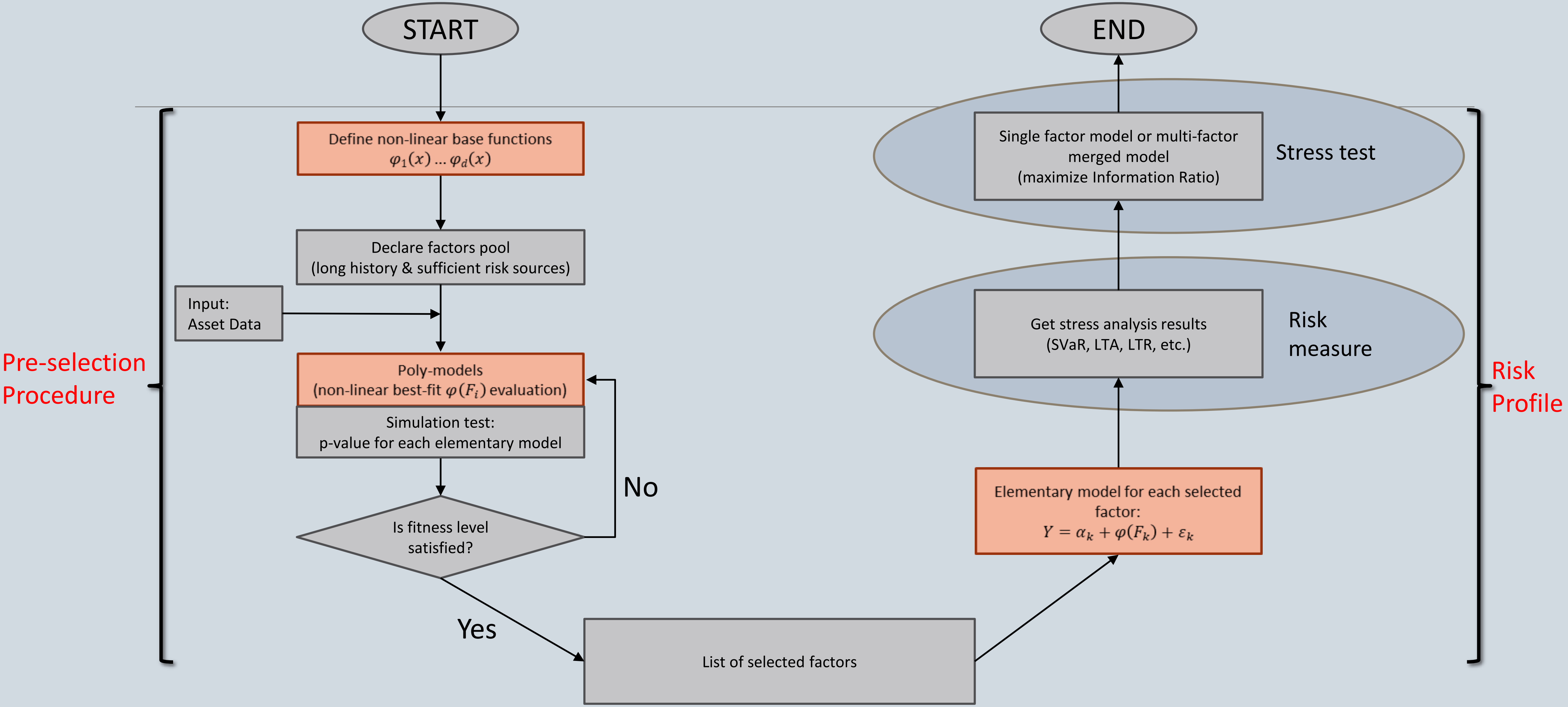
Input:
Factor Pool F_j

Output: Poly-model =
List of Single-Factor Models
Results in:
$\Rightarrow$ Risk Measure: Stress VaR
$\Rightarrow$ Dominant Risk Factors
$\Rightarrow$ Stress Tests

Explore the Black Box



Poly-model Algorithm Structure Example



Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Polymodels: The Math for Dominant Factor™ Analysis

Multi-Factor Model

$$Fund = \lambda_1 Fact_1 + \cdots + \lambda_n Fact_n + \alpha$$

- Coefficient λ_i are fixed
- Factor set $\{Fact_1, \ldots, Fact_n\}$ is frozen

Polymodel: Collection of Models:

- Linear: $Fund = \beta_i Fact_i + \alpha_i \quad i = 1 \ldots n$
- Nonlinear: $Fund = \varphi_i(Fact_i) + \alpha_i \quad i = 1 \ldots n$

  (one can extend the formula with lagged terms)
- Score each model by relevance in ***extreme*** scenarios

  Score = –Ln(*p*-value)
- *p*-value = Probability that, given the data, the observed relation is spurious

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Polymodels: Relation with Multi-factor Models

Linear case

- $Fund = \beta_i Fact_i + \alpha_i i = 1 \dots n$
- $Fund = \lambda_1 Fact_1 + \cdots + \lambda_n Fact_n + \alpha$
- $Cov(Fund, Fact_i) = \beta_i Var(Fact_i) = \sum_{j=1}^{n} \lambda_j Cov(Fact_i, Fact_j)$
- $(\lambda_1, \dots, \lambda_n) = Cov(Fact)^{-1}(\beta_1 Var(Fact_1), \dots, \beta_n Var(Fact_n))$
- The uncertainty on λ_l's depends on colinearity of factors
- Badly conditioned Cov matrix $\Rightarrow$ Low Information Ratio

Nonlinear Modelling

- Hermitte Polynomials H_k:

$$\varphi_i(Fact_i) = \sum_{k=0}^{d} H_k(Fact_i) + \alpha_i$$

- Nonlinear Multi-factor model by inverting $Cov(H_k(Fact_i))$
- Improve Information Ratio with LOESS Regression

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Relation between Multi-Factor and Polymodels

Theorem (Cherny-Douady)

Given $X_1,\dots,X_n$ and functions $\varphi_1,\dots,\varphi_n$ such that $E[\varphi_1(X_1)] =\dots= E[\varphi_n(X_n)]$

Does there exist a function φ of n variables such that:

$$\forall i \qquad E[\varphi(X_1,\dots,X_n)\,|\,X_i] = \varphi_i(X_i)$$

Answer: the minimum variance solution is given by

$$\varphi(X_1,\dots,X_n) = \psi_1(X_1) +\dots+ \psi_n(X_n)$$

It exists provided an ellipticity condition is satisfied: for any $(\varphi_1,\dots,\varphi_n) \in L^2(X_1)$ x…x $L^2(X_n)$

$$\left\| \sum_{i=1}^{n} \varphi_i(X_i) \right\|_{L^2}^2 \geq c \sum_{i=1}^{n} \left\| \varphi_i(X_i) \right\|_{L^2}^2$$

Dominant Factor™ Analysis: Stress Testing

Model Selection

- For each subset of indices $I = (i_1,\ldots,i_q)$, merge models as above
- Compute the Information Ratio:

  IR = Merged Impact / Uncertainty

- Find the subset I with the highest Information Ratio

Stepwise Regression

- Find the factor i_1 with highest Information Ratio
- Take this factor as given. Find the second factor i_2 such as, jointly with i_1, the Information Ratio is maximum
- Repeat until the Information Ratio cannot be increased
- Try to remove factors while increasing the Information Ratio
- Stop when it is not possible to add – or remove – factors

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor™ Analysis: Information Ratio

Given $I = (i_1, \ldots, i_q)$ and factor stress values $\left(x_{i_1}, \ldots, x_{i_q}\right)$ we compute the joint impact by merging single factor models:

$$Impact = \sum_{i \in I} \lambda_i^k H_k(x_i)$$

where λ_i^k are the coefficients of the merged multi-factor nonlinear model.

The uncertainty of the estimate is given by the covariance matrix of coefficients $\left(\lambda_i^k, \alpha_I\right)$, which can be redeemed from the Fischer information matrix.

$$IR = \frac{Expected\ Impact - Unconditional\ Fund\ Expectation}{Std\ Dev(Uncertainty\ of\ Impact)}$$

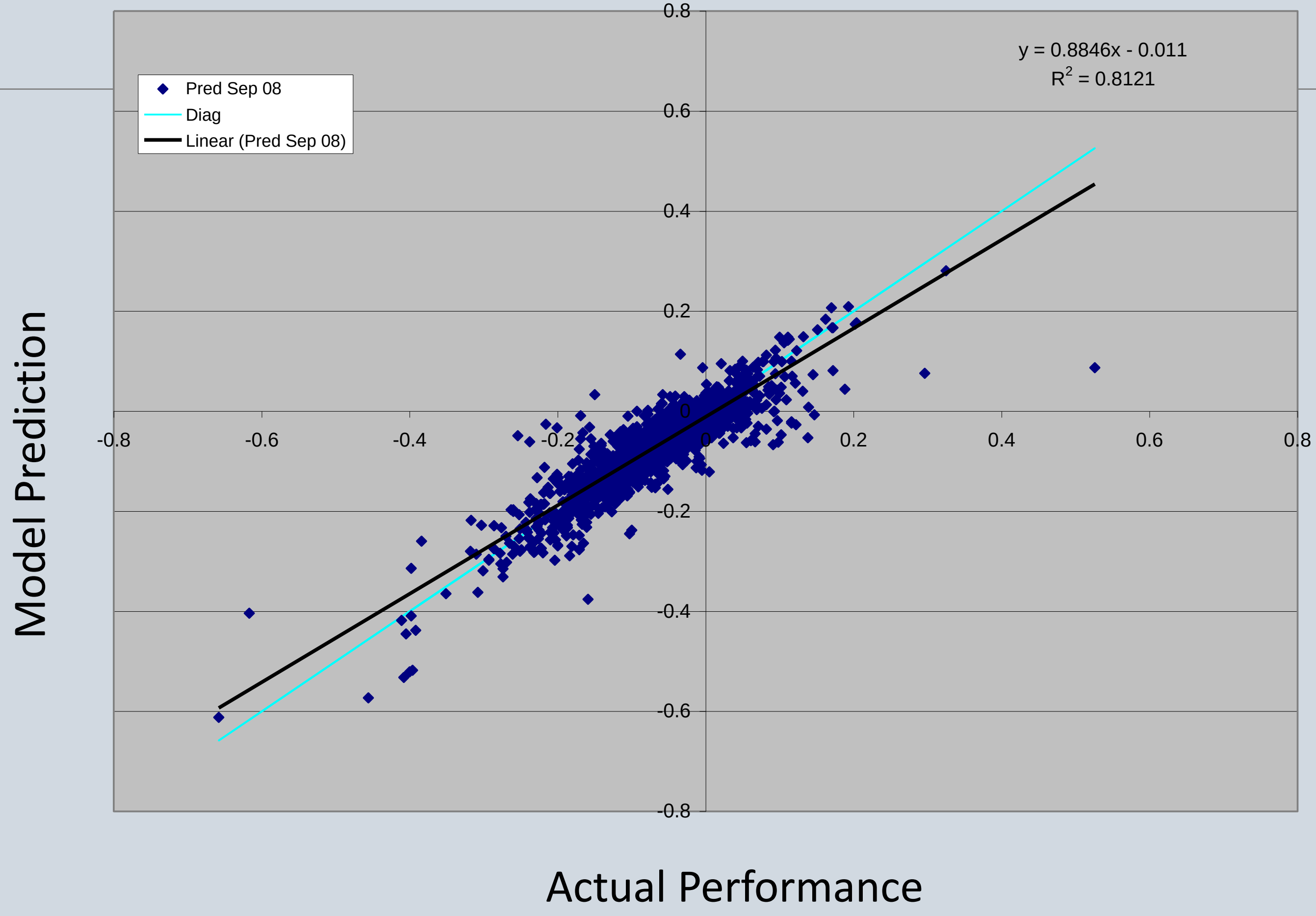
Account for *small sample bias* and non-Gaussian input distributions

p-value = Percentile of $\mathrm{E}(Fund)$ in the distribution of Impact

LOESS Regression: Weighted linear model $\Rightarrow$ Better Information Ratio

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor™ Analysis: Prediction Capability



Dominant Factor Analysis
using ~200 Risk Factors
Adaptive *p*-value computation
Most relevant factor only

3000+ fund returns in Sep 08
Coefficients calibrated on [Sep 05, Aug 08]

Data from HFR

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor™ Analysis: Stress VaR

Stress VaR is a risk measure that combines stress tests and value-at-risk.

It relies on "poly-models" for the estimation of the distribution of future returns.

It is generated from market histories that include past crises, and draws on a sufficient volume of factors, so as to account for all possible risk sources.

Nonlinear models capture extreme risks – in particular, the impact of liquidity gaps.

Therefore, the Stress VaR unveils hidden risks by identifying drivers of returns.

For each *Dominant* Factor f_k, estimate 99% VaR using (very) long-term history: V_k

- Use "shadow estimate" by fitting a fat-tailed distribution on historical returns

Check the impact of this scenario on the fund, using the relevant model: $\varphi_k(V_k)$

Gaussian estimate of residual risks for this particular model: R_k

$$\text{StressVaR}(f_k) = \sqrt{\varphi_k(V_k)^2 + R_k^2} \qquad \text{StressVaR} = \text{Max}_{\text{Dominant } fk}\ \text{StressVaR}(f_k)$$

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC
54

Risk Measurement by Polymodels

Handle 100’s of Market Factors

Model **rare events** (“Black Swans”)

More accurate when needed, than when not needed!

- ***Tail concentration*** effect

Suited for **risk measurement** *and* **stress scenarios**

- Prediction from individual factors can be merged
- Risk measure = STRESS VaR (worst case) includes ***hidden risks***

Can be aggregated for a portfolio

- Risk contributions involve ***extreme correlations***
- Superior allocation and optimization

Dominant Factor Long-Term Alpha

Poly-models are a "translator" between factors distribution and fund returns distribution

Funds lack long history, but factors went through many crises

Estimate the *long-term distribution* $f_i(x)dx$ of each dominant factor X_i

- Estimation should span a full economic cycle (e.g. 40 years) in order to handle all phases

Deduce the *long-term distribution* of fund returns according to X_i

$$g_i(y)dy = \frac{f_i(\varphi_i^{-1}(y))}{\varphi_i'(\varphi_i^{-1}(y))}dy$$

Merge distributions $g_i(y)dy$ according to their relevance into $g(y)dy$

$$\text{Long}-\text{term Alpha} = E_g(Y)$$

$$\text{Long}-\text{term Risk}-\text{adjusted Expected Return} = \frac{\text{Long}-\text{Term Alpha}}{\text{Stress VaR}}$$

Long-Term Alpha & Fund Ranking

Long-Term Alpha converts optionality into expected returns

- Gamma ч Theta

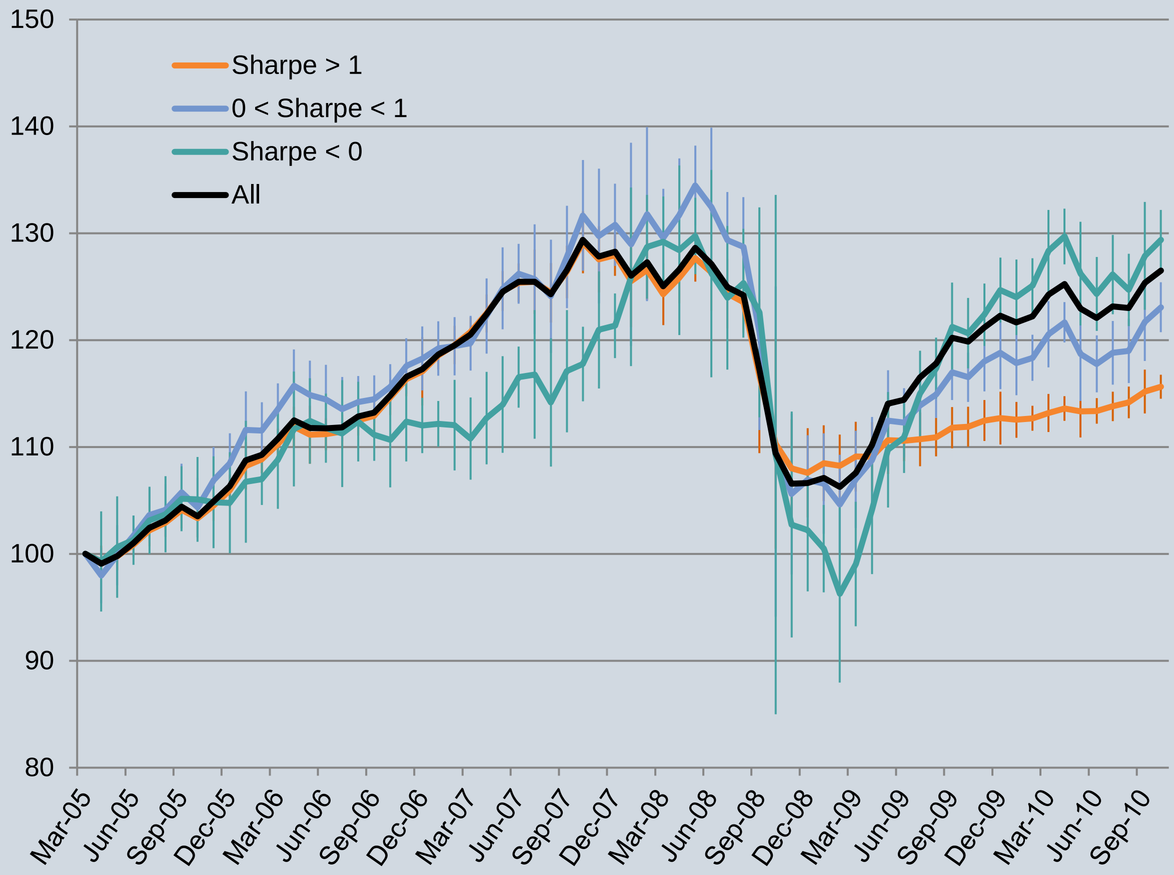
If correctly estimated, it is the same across all factors

- Powerful and robust predictor of long-term returns

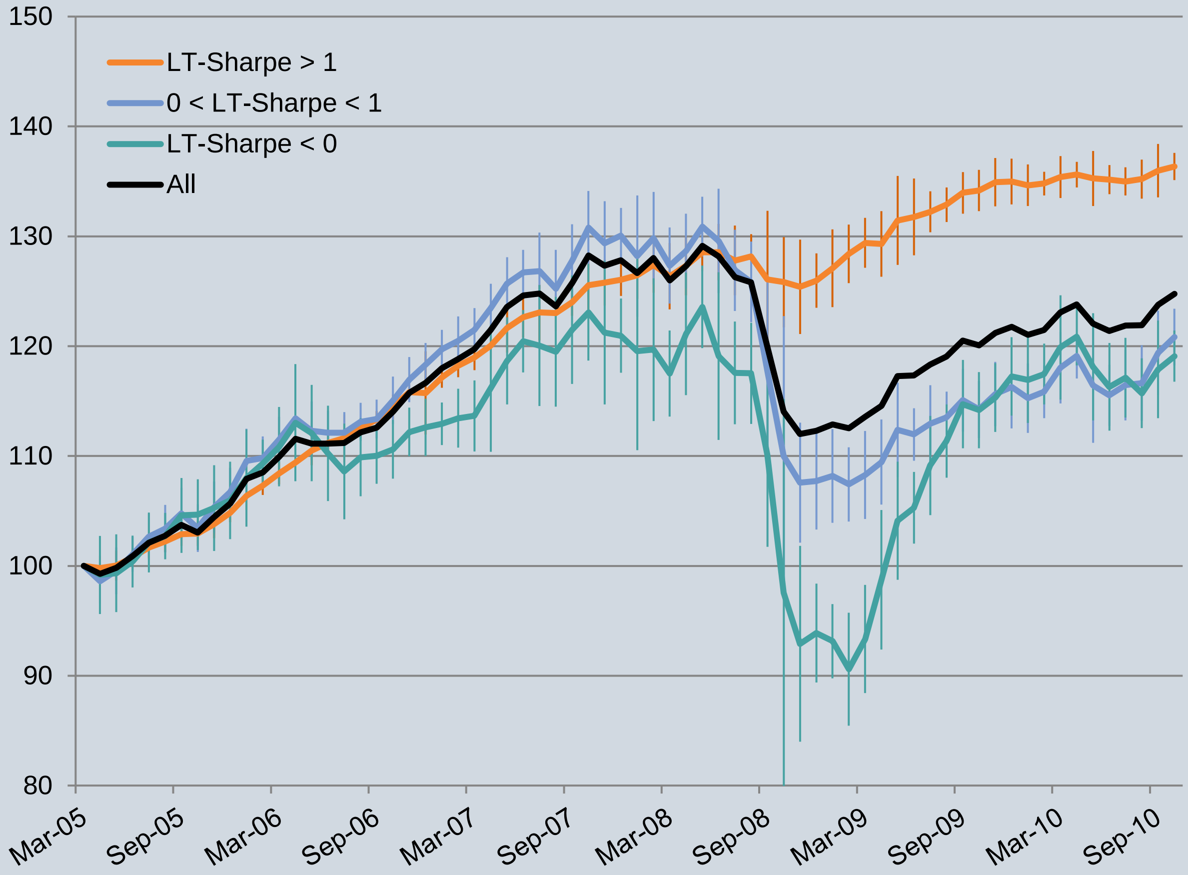
Back-test Dominant Factor亻 Analysis:

- Compute LT-Alpha and StressVaR of each fund at times $t_1,\ldots,t_n$
- Sort funds by LT-Alpha bucket (or LT-RAER bucket)
- Allocate by *extreme risk parity*: $\boldsymbol{weight \sim \frac{1}{Stress\ VaR^{\delta}}}$
- Tested over Jan 2005 – Oct 2010 on 500 hedge funds (HFR)
  - Dominant Factor亻 Analysis computed on 36 months, updated every 6 months
  - 4 months delay before investment
- Comparison with 36M performance and volatility

Hedge Fund Portfolio: Sharpe Ratio Selection Iso-Volatility Allocation



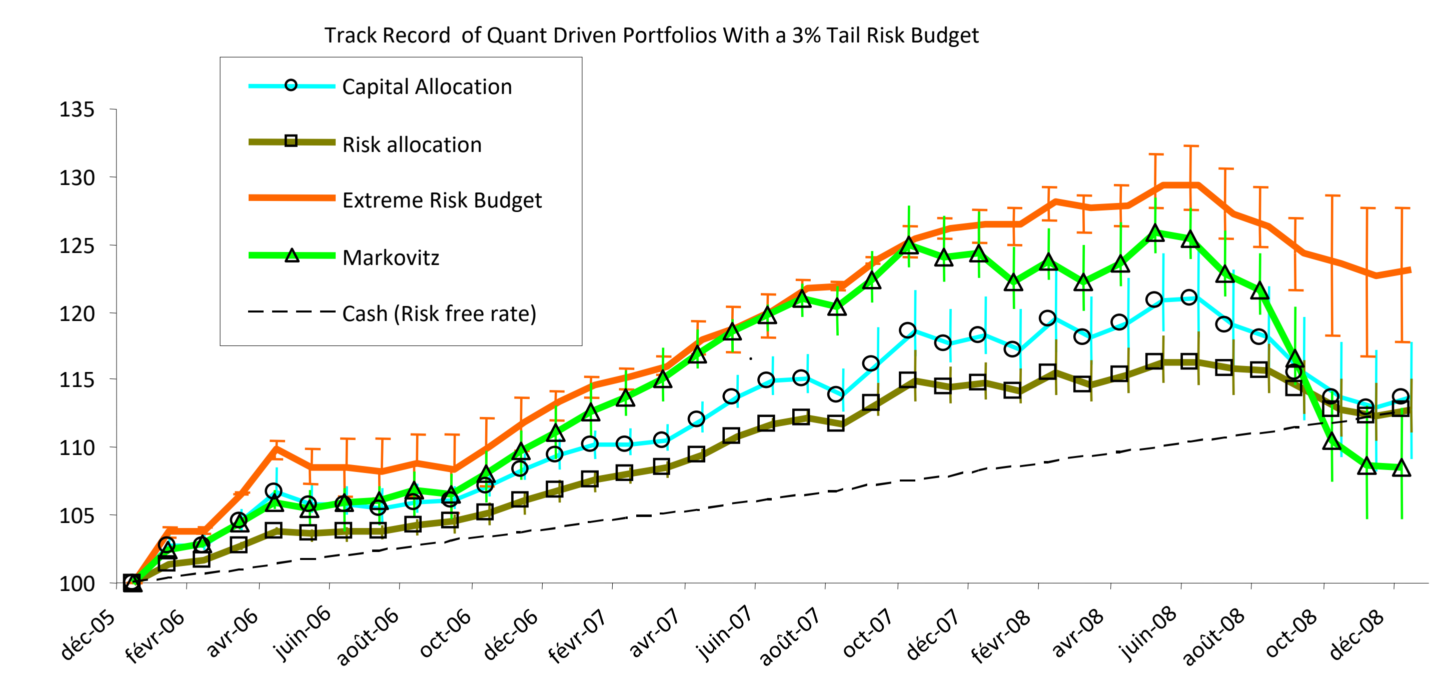
Dominant Factorイ Analysis on Hedge Funds: LT-RAER Selection Iso-StressVaR Allocation



Pay for "Hedging Costs", Relaxing "Business as Usual" Risk Constraints

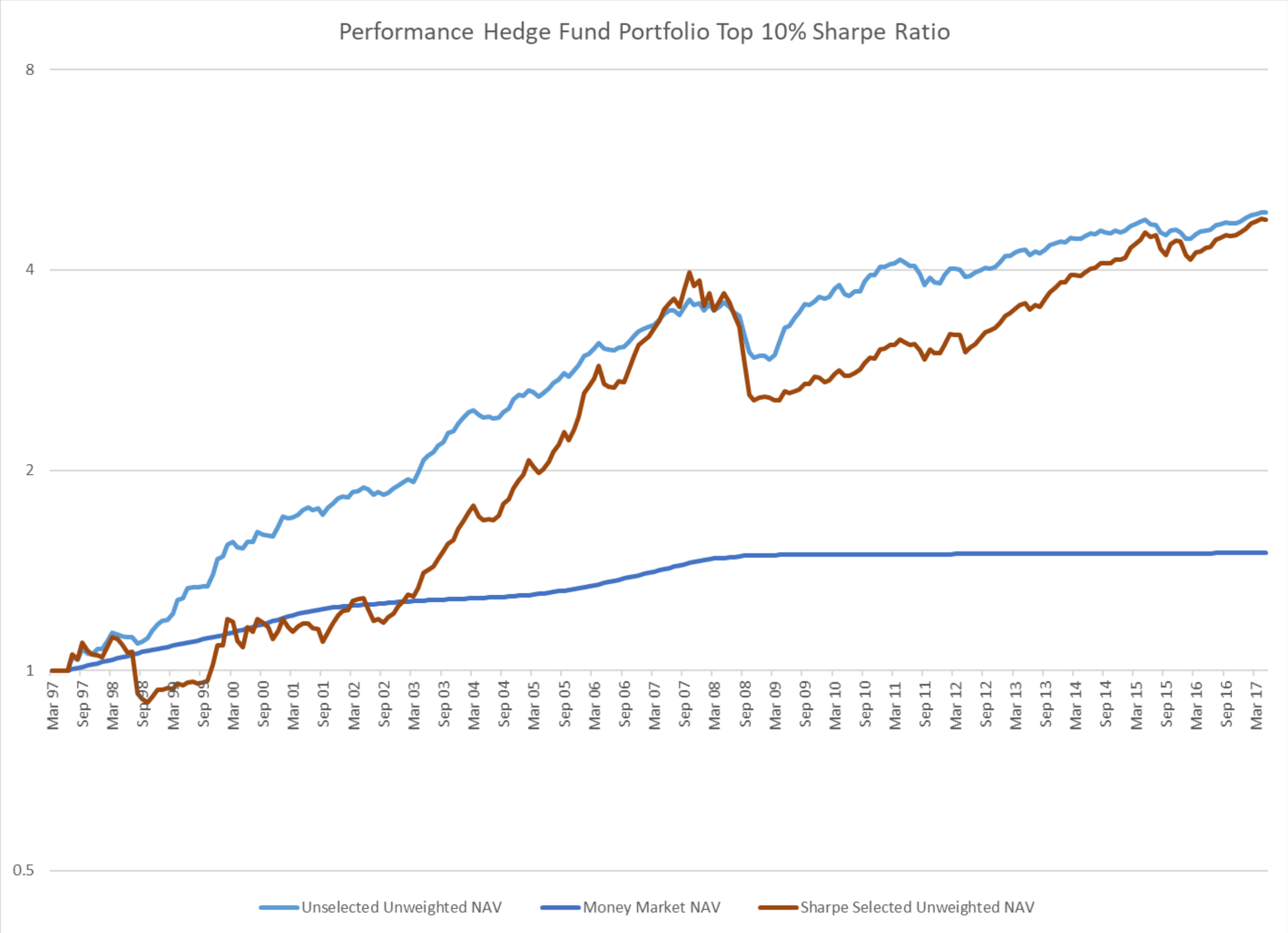
An additional risk premium can be extracted by relaxing "normal" risk constraints. Extreme risk budgeted portfolios, using Dominant Factor™ analysis, outperform Markovitz under business-as-usual conditions – despite exhibiting a lower Sharpe ratio, and simply because they have higher volatility. In troubled times, Markovitz collapses while Dominant Factor™ analysis holds up well.

Source: HFR Hedge Fund Database



Capital Allocation	Medium Performance	Medium Risk Control
Risk Allocation	Weak Performance	Risk Control OK
Markowitz	Performance OK	Bad Risk Control
Extreme Risk Budget	Performance OK	Good Risk Control

Hedge Fund Index by Ranking: Sharpe

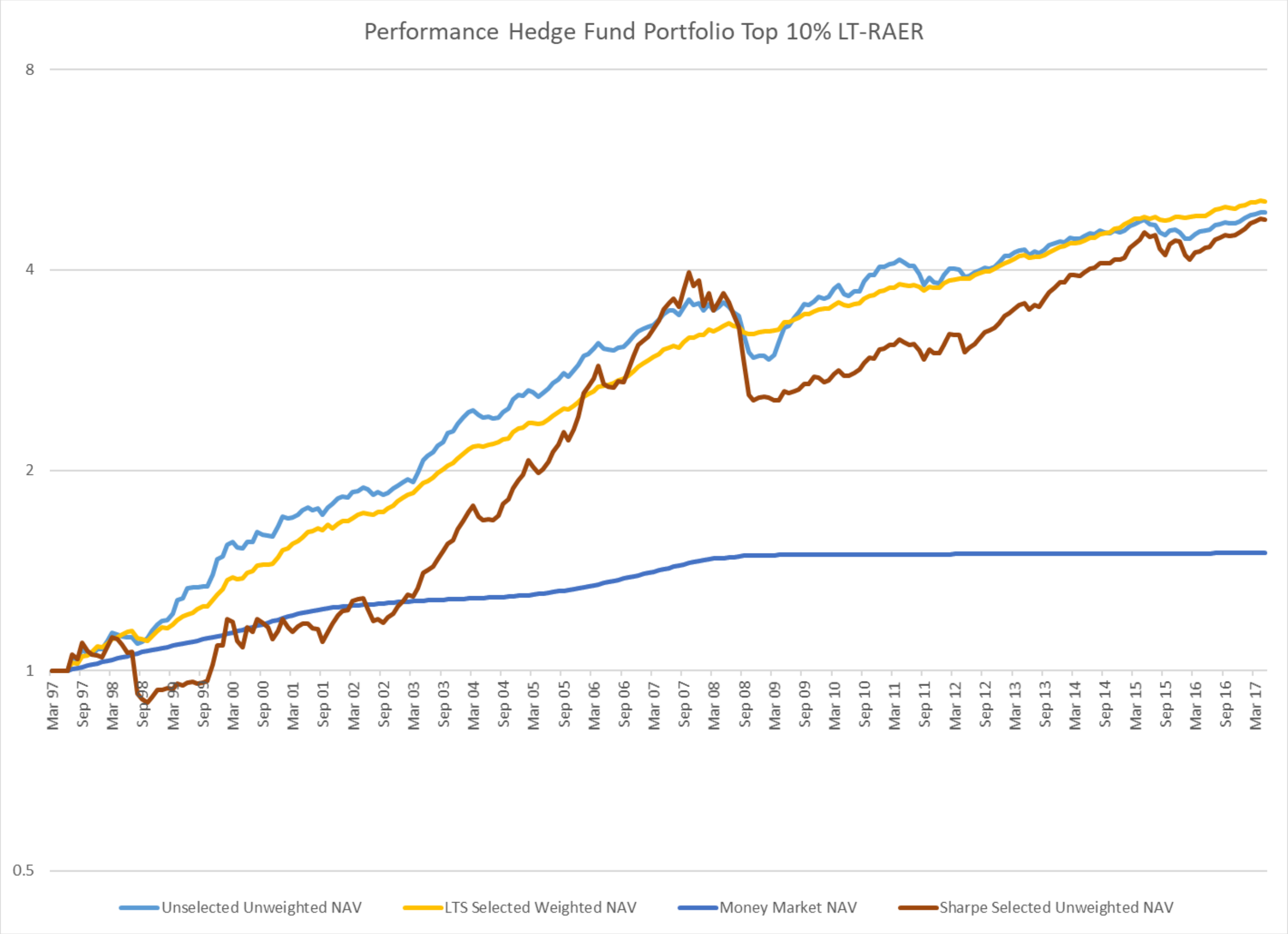


Performance of a monthly rebalanced portfolio of the top 10% hedge funds by Sharpe Ratio vs. the equally weighted portfolio of all funds

3 months lag to account for redemption delays.

The best funds by Sharpe Ratio are also the most fragile ones

Hedge Fund Index by Ranking: LT-RAER

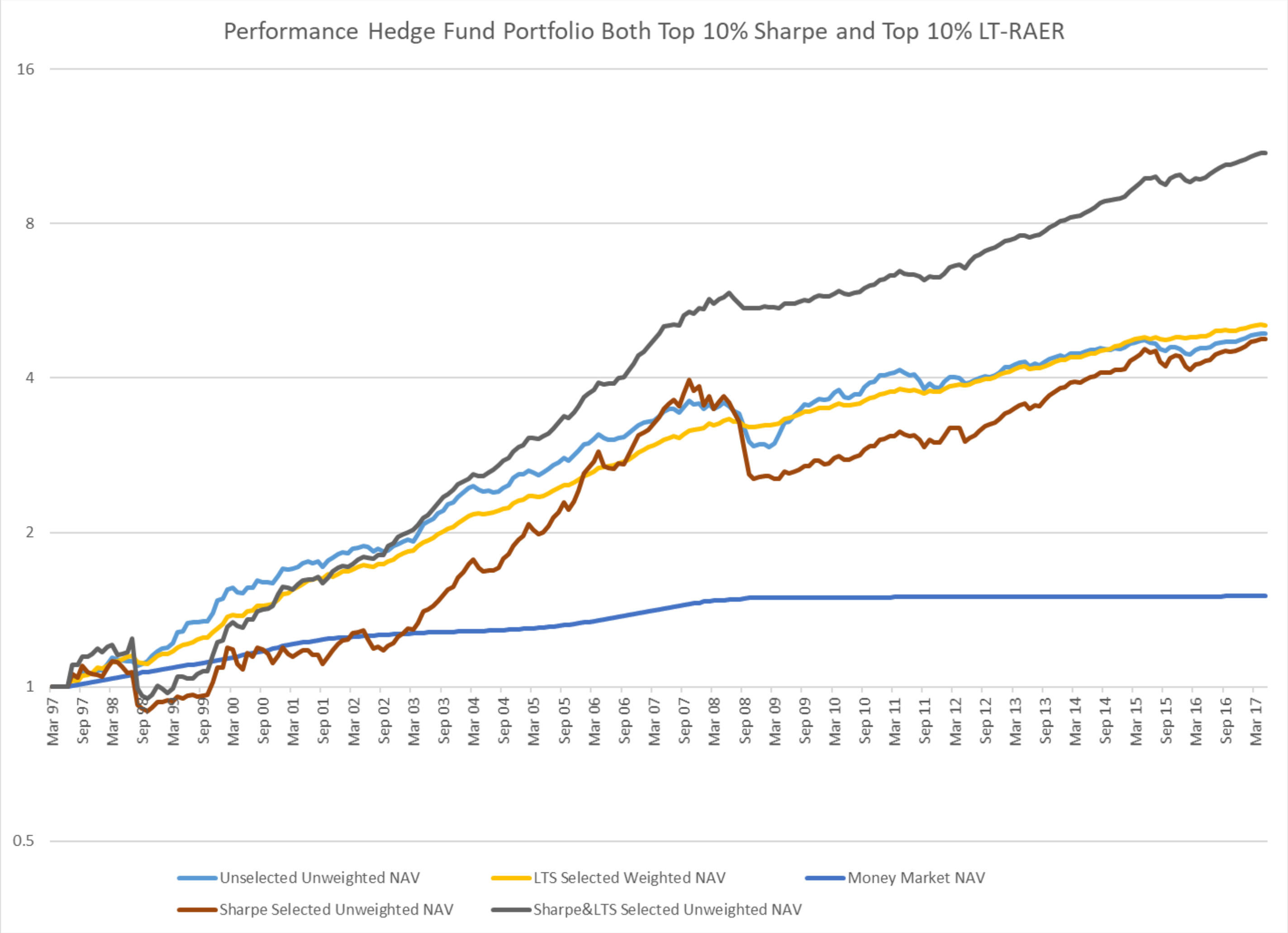


Performance of a monthly rebalanced portfolio of the top 10% hedge funds by Long-Term Risk-Adjusted Expected Return (LT-RAER) vs. Sharpe Ratio

3 months lag to account for redemption delays.

The best funds by LT-RAER resist the crisis, while keeping the overall performance.

Hedge Fund Index by Ranking: LT-RAER + Sharpe

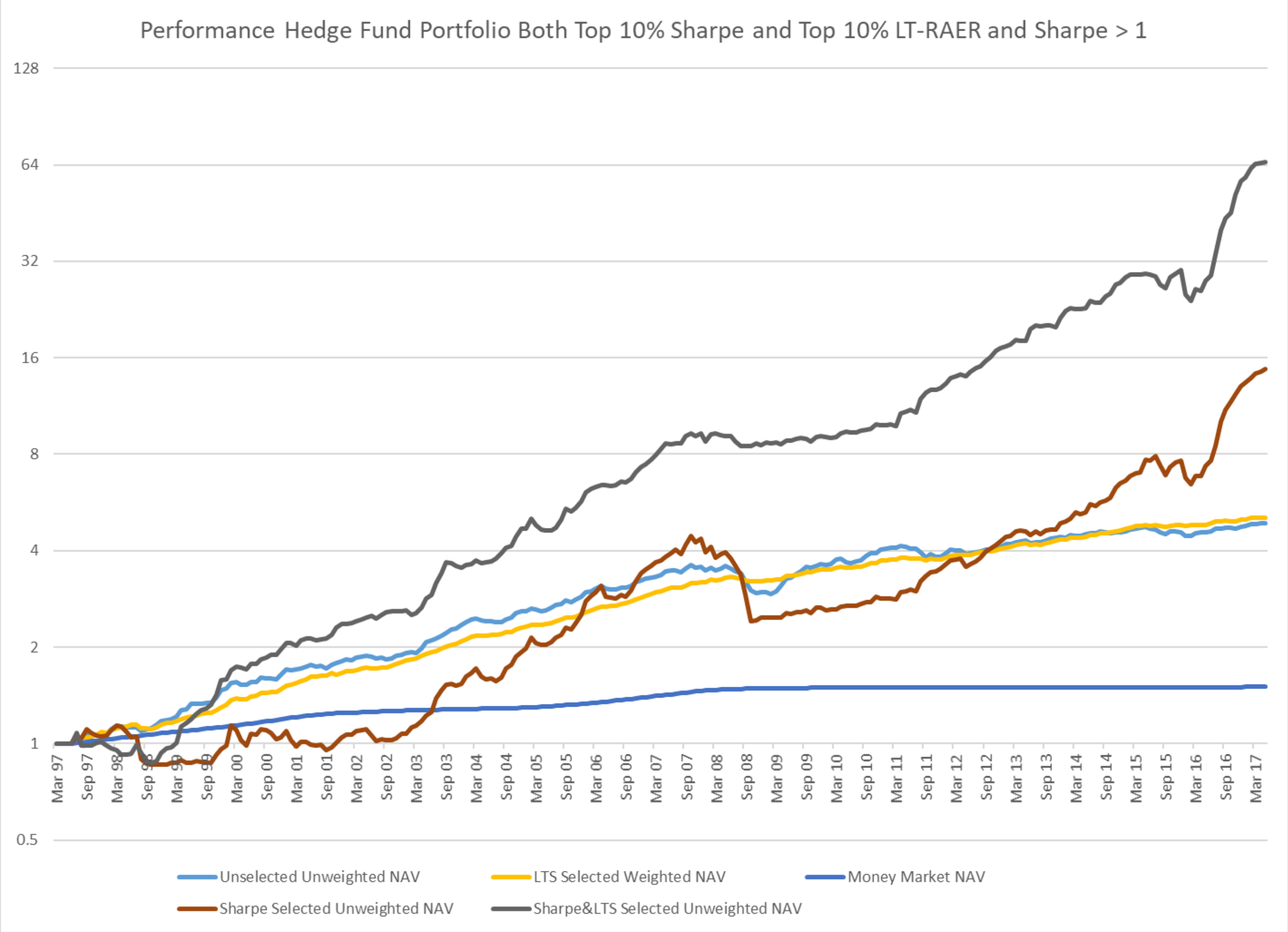


Performance of a monthly rebalanced portfolio of hedge funds that are in the top 10%, both by Long-Term Risk-Adjusted Expected Return (LT-RAER) and by Sharpe Ratio

3 months lag to account for redemption delays.

In market rally, the portfolio follows the market, but LT-RAER protects the downside in the down turn.

Hedge Fund Index by Ranking: LT-RAER + Sharpe



Performance of a monthly rebalanced portfolio of hedge funds that are in the top 10%, both by Long-Term Risk-Adjusted Expected Return (LT-RAER) and by Sharpe Ratio, with Sharpe > 1.

3 months lag to account for redemption delays.

In market rally, the portfolio fully captures the upside, but LT-RAER protects the downside in the down turn.

Dominant Factor™ Analysis: Fund Categorization

Categorize by Assets, Geography and Trading Strategy

- Most databases
- Upon manager declaration. May drift…

Categorize by Correlation and Clustering

- Very Unstable
- Not always meaningful (e.g. Macro manager switches exposure…)

Categorize by proximity of Polymodel, then clustering

- Much more stable
- Pros and cons of pure statistics: avoids being fooled, but can be lured…
- Benchmark often identified as the most dominant factor (if in the factor set)

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC
65

Dominant Factor™ Analysis: Supervised Learning

Question: Fund risks and future returns?

Input: Polymodel (instead of series of returns)

- Uses market information $\Rightarrow$ Sensitivity to risk factors

Answer:

- Selects funds that have a *convex* (antifragile) response to dominant factors
- Rejects funds that have a *concave* (fragile) response to dominant factors

Advanced methods (to be tested)

- Deep learning
- Adversarial learning

Learning to "guess the future": contradicts Market Efficiency Hypothesis

Learning risks with behavior under stress: Use sensitivity to factors as input

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor™ Analysis: Categorization, Benchmark

Question: How to sort out the various funds? Which Benchmark

Input: Polymodel (instead of series of returns)

- Uses market information $\Rightarrow$ Sensitivity to risk factors

Method: Define a "distance" between funds

- Weighted average of "distance" between one-factor models, weighted by the product of the scores of the funds
- Clustering of the set of funds using this distance.
- Benchmark = Cluster average

Answer:

- 80% in line with traditional categories, but some surprises
- Cluster averages are "purer", more stable than traditional indices based on strategy
- When a fund departs its "benchmark" (cluster average), a true style drift is observed

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor™ Analysis

Handle 100's of Market Factors

Model **rare events** ("Black Swans")

More accurate when needed, than when not needed!

- ***Tail concentration*** effect

Suited for **risk measurement** *and* **stress scenarios**

- Prediction from individual factors can be merged
- Risk measure = STRESS VaR (worst case) includes ***hidden risks***

Can be aggregated for a portfolio

- Risk contributions involve ***extreme correlations***
- Superior allocation and optimization

Dominant Factor™ is a protected Trademark of DataCore Innovations, LLC (2015) and had been used with the permission of DataCore Innovations, LLC

Dominant Factor≺ Analysis: One of the Most Powerful Buy-side Quant Tool

- Dominant Factor≺ as a Risk tool
  - Stress VaR
  - Single Factor Stress Testing with Single factor model
  - Full Scenario Testing with Max Information Ratio
- Dominant Factor≺ as a Replication and Stress Testing tool
  - Multi-model selection with Max Information Ratio
- Dominant Factor≺ as a Selection Tool
  - Long-term Risk-adjusted Return
- Dominant Factor≺ as Allocation Tool
  - Extreme Risk-parity
  - Extreme Risk Budgeting

Building a Data Set

Data Source Quality:
- Audited or self contribution?
- Appraisal method? Can be “sound” but statistically invalid (e.g. conservative)!
- Completeness? Poor completion algorithm…
- Corrections for security transactions

Investability
- Is everything reflected in data? Fees, defaults, dividends, etc.
- Survivorship bias: how was the market at the date?
- Feasibility of investment, redemption? Gates?
- Market depth: volume of transactions, market impact?

Computer and Statistical Tractability
- Homogeneity
- Completeness
- “Signal” is meaningful? Noise? Approximation? Digitization?

Building a Test Bench

Set of funds for test

- Must be as faithful a representation of the investment context as possible
- No overlap between the dates or set of funds used for calibration and learning and for testing
- Include artificial noise and check robustness of results

Factor Set Selection

- Are selected factors good for prediction? $\Rightarrow$ P-square
- A factor can be very good for prediction, even if not in portfolios (e.g. Russian market index for small caps, Australia and Canada for CTAs)
- Use AI to sort out good and bad factors

Portfolio Construction

Markowitz:

Maximize Sharpe Ratio

Black-Litterman:

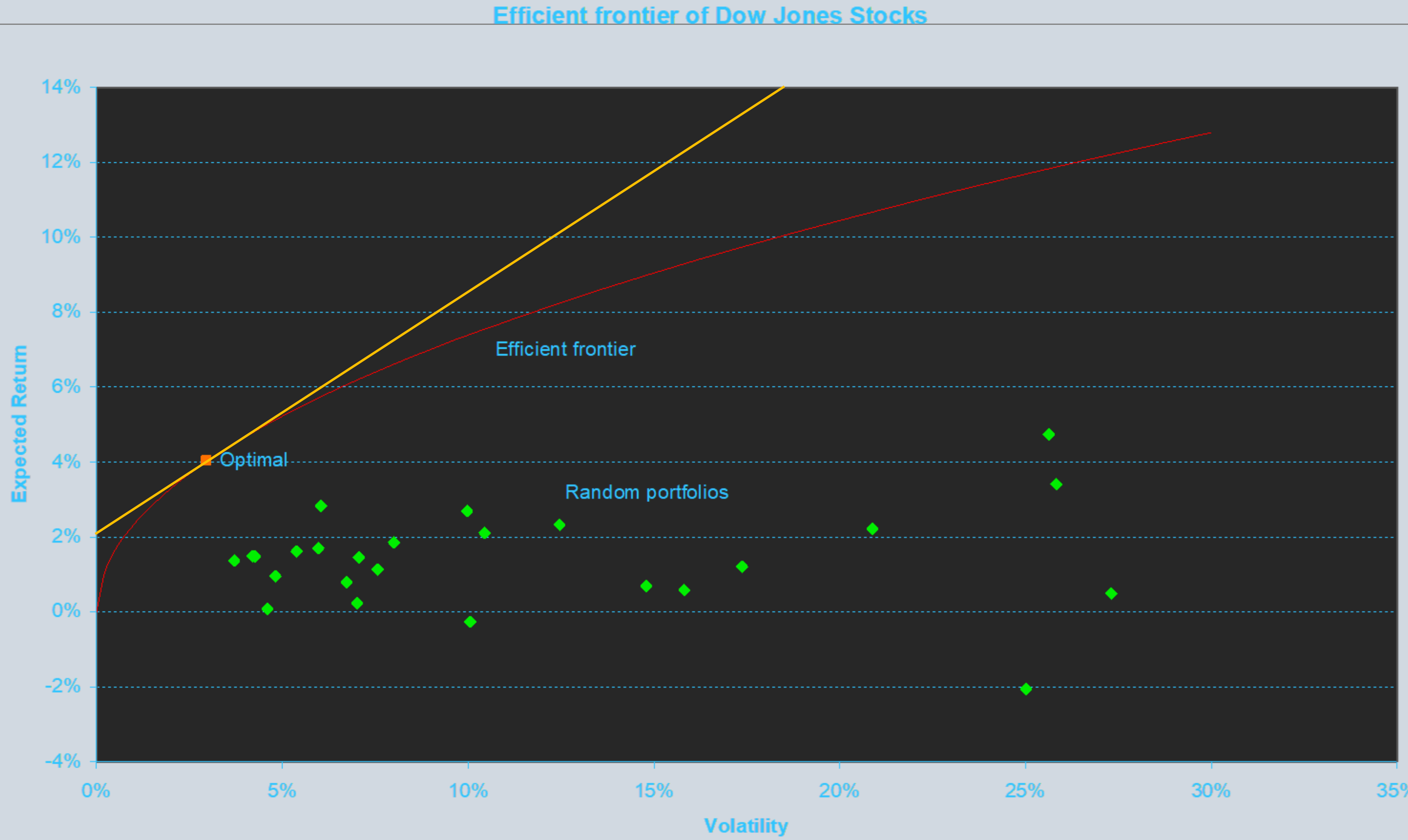
Bayesian Statistics

Key feature:

Contributions to Risks and to Expected Returns are Proportional

Problem:

Relies on Correlations…



Portfolio Construction

Extreme risks: Diversification vanishes

- Correlations surge

Extreme risks must de evaluated through sensitivity to risk factors and stress testing

- Check diversification under each stress test $\Rightarrow$ Risk contributions
- Include pre-existing and known risks

Optimal Portfolio

Risk Contributions ~ Return Contributions

Machine learning:

- Get more relevant risk contributions, based on better understanding of behavior under stress
- Design Macro-hedge against main risks or pre-existing ones (e.g. energy)

Aggregation of Polymodels

- Aggregation of Elementary Models

$$P = \sum_{k=1}^{n} w_k S_k \quad \Rightarrow \quad E(P|X) = \sum_{k=1}^{n} w_k E(S_k|X)$$

- Aggregated Portfolio Score

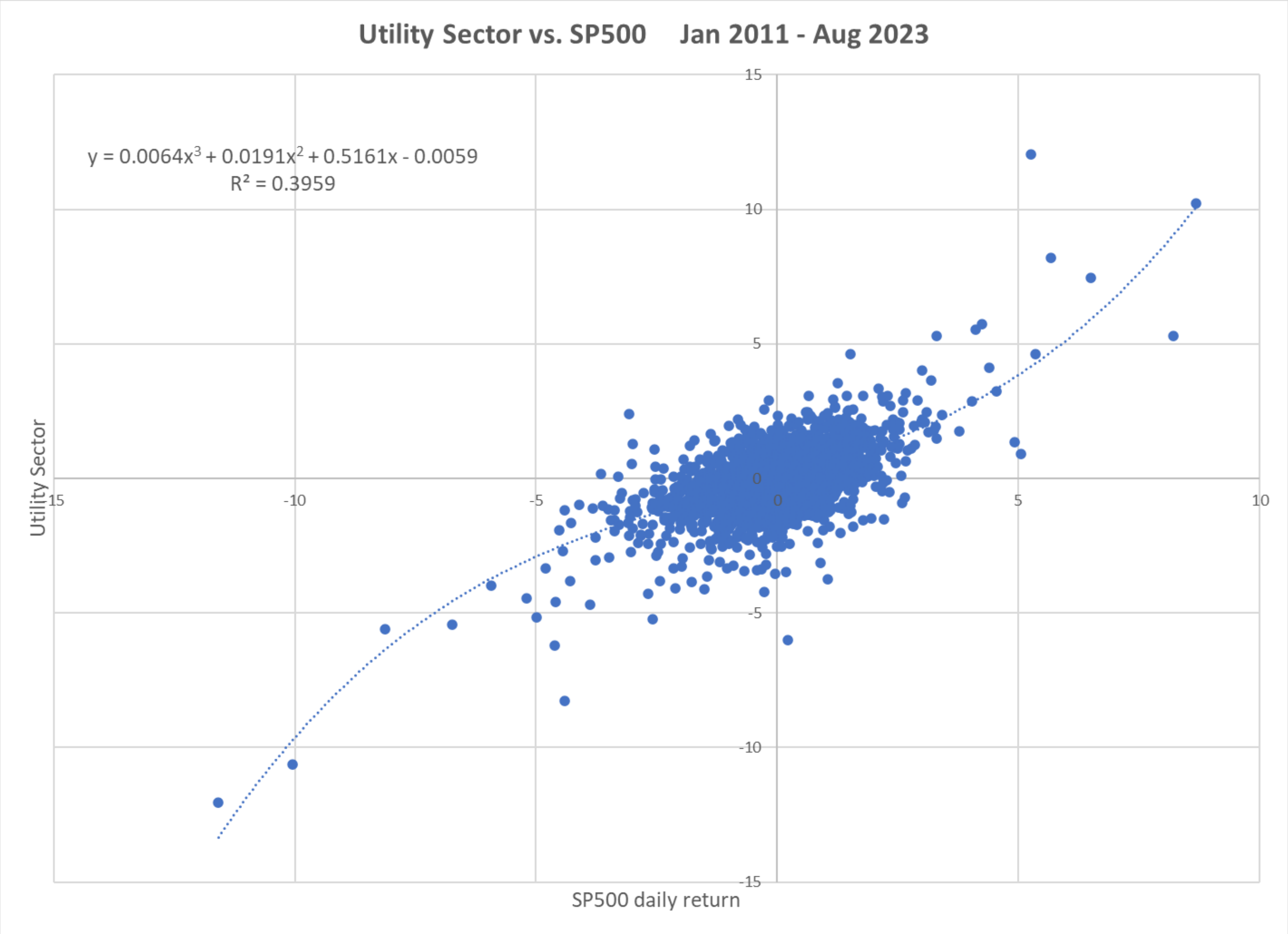
$$S_k = E(S_k|X) + \varepsilon_k$$

- Compute a proxy of the score $-\mathrm{Ln}(p\mathrm{-value}(S_k, X))$ using only $R\mathrm{-square} = 1 - \frac{\mathrm{var}\varepsilon_k}{\mathrm{var}S_k}$
- Compute an implied covariance matrix of the various ε_k
- Compute the resulting aggregated R-square
- Compute the implied score $-\mathrm{Ln}(p\mathrm{-value}(P, X))$

Extreme Risk Budgeting

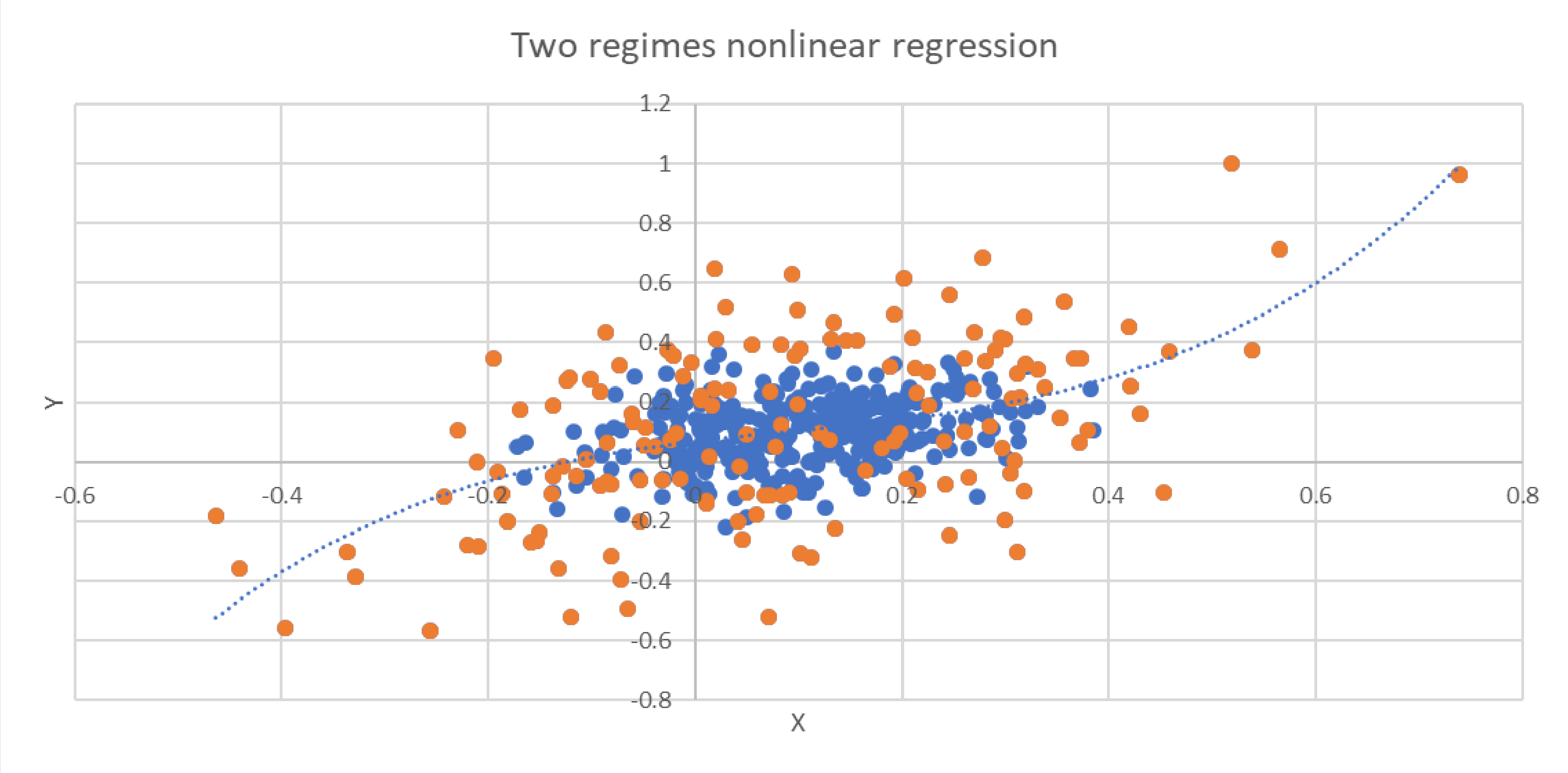
- ❑ Budget for the next crisis to secure long-term returns
- ❑ During extreme market conditions, monitoring credit and liquidity risk ⬄ **Hidden** Market Risks
- ❑ **Polymodels** and **Stress VaR** unveil "hidden risks" by including long-term **market history** into **risk analysis**, focusing on **extremes**
- ❑ Measuring "hidden" market risk means integrating **gamma**, **long-term factor risk** and return **smoothing**
- ❑ Monitoring "hidden" market risk budget implies **shifting from static allocations to stable risk budgets** per factors, thereby reflecting ALM constraints & long-term views
- ❑ These parameters help discriminate between "lucky" managers (who generate returns based on hidden risks), and **real talented ones**!

Hidden Markov Model in US Equities

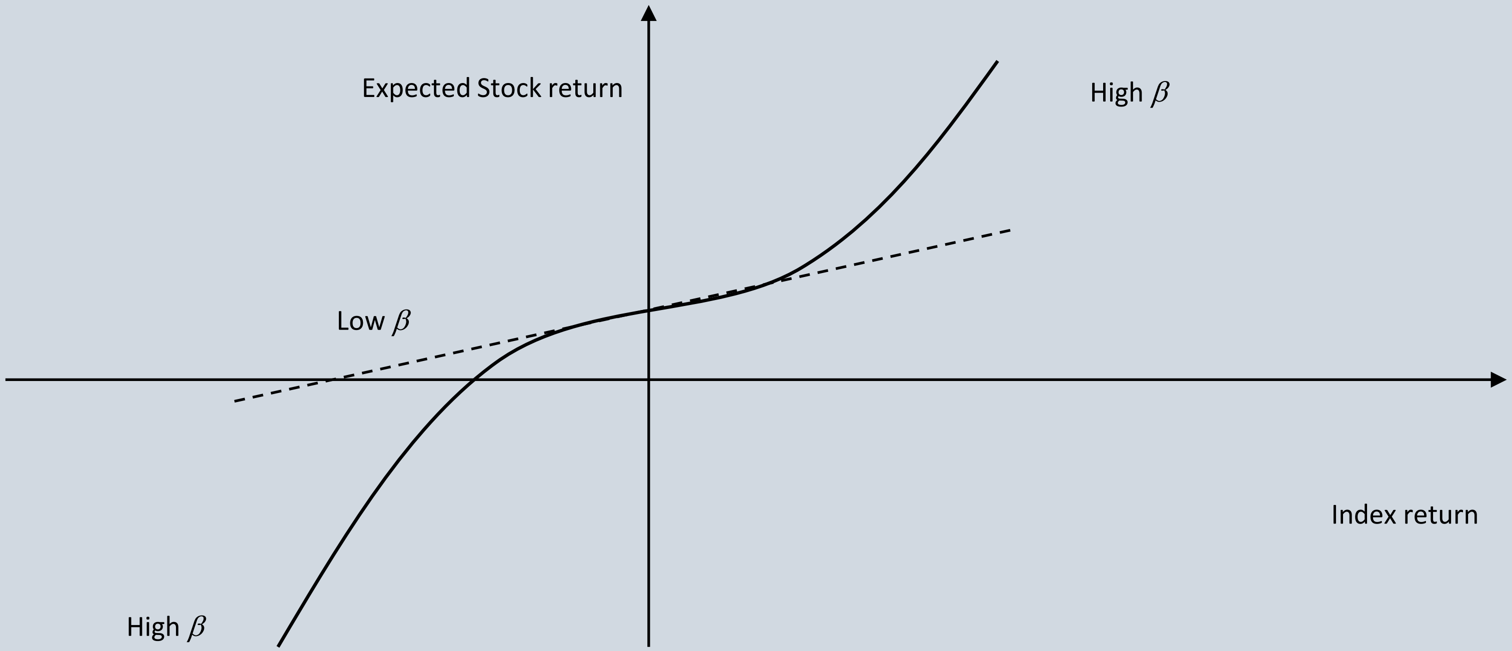


Polymodels in a Hidden Markov Model

	Regime 1	Regime 2
Proba	69%	31%
Mu X	10%	10%
Mu Y	10%	10%
Sigma X	10%	20%
Sigma Y	10%	30%
Beta(Y/X)	0.3	0.8
Specif Y	9.5%	25.4%



Polymodels in a Hidden Markov Model



Polymodels in a Hidden Markov Model

Risk Factor: $X(t)$

Invested Asset: $Y(t)$

Bayesian Computation of $E(Y|X)$

- Prior regime probabilities: ergodic limit π_k (uninformed Jeffreys prior)
- Given the value of X, compute the implied (posterior) probabilities $p_k(X)$
- $E(Y|X) = \sum_{k=1}^{m} p_k(X) E_{R_k}(Y|X)$
- $E_{R_k}(Y|X) = a_k X + b_k \Longrightarrow E(Y|X) = f(X)$ with nonlinear f
- The more probabilities are sensitive to X, the more nonlinear is $f \Rightarrow$ Crisis indicator

HMM is a *dynamic model* that can be simulated

Polymodels are a *statistical analysis* of the dynamics

Bayesian Estimation of Polymodel

- Prior probability $\pi_0(R)$ (ergodic limit or given by the Markov chain)
- For a given factor shift x, the posterior probability $\pi_1(R,x) \sim \pi_0(R)\ \varphi_R(x)$

$$\pi_1(R,x) = \frac{\pi_0(R)\varphi_R(x)}{\sum_R \pi_0(R)\varphi_R(x)}$$

$$\mathbb{E}[Y|X=x] = \frac{\sum_R \pi_0(R)\varphi_R(x)\mathbb{E}_R[Y|X=x]}{\sum_R \pi_0(R)\varphi_R(x)}$$

- Example: Each regime is Gaussian

$$\mathbb{E}_R[Y|X=x] = \alpha_R + \beta_R x$$

$$\varphi_R(x) = \frac{1}{\sigma_R\sqrt{2\pi}} e^{-\frac{(x-\mu_R)^2}{2\sigma_R^2}}$$

Polymodels to Identify and Calibrate Regimes

Given a Candidate List of Regimes and Regime Switching Probabilities

- Compute Polymodels of a series of indices

Separately estimate the Polymodel of these indices from historical data

Estimate the Likelihood of the Candidate list

Find the candidate list that Maximizes the Likelihood

No clustering algorithm $\Rightarrow$ Robustness of results

Can be reproduced each month $\Rightarrow$ Adapts to current market situations

Research in progress…

Fees: Fair or Too expensive?

What is the Benchmark?

Easiness and Cost of Benchmark?

Alpha vs. Extreme Risk of Fund vs. Benchmark

- Is it "True Alpha" or "Theta of an Option"?
- Only the portion of performances above extreme risk should produce fees

Are there equivalent, yet cheaper, products?

- Compare polymodels to find similar products
- Correlation is not enough and doesn't account for extreme risks

Build a Fee Table by fund category (in progress…)

Skill vs. Luck: Simulations and Actual Trading

Simulated performances may show skills.

Actual success can be due to luck!

Identify performance drivers and hidden risks

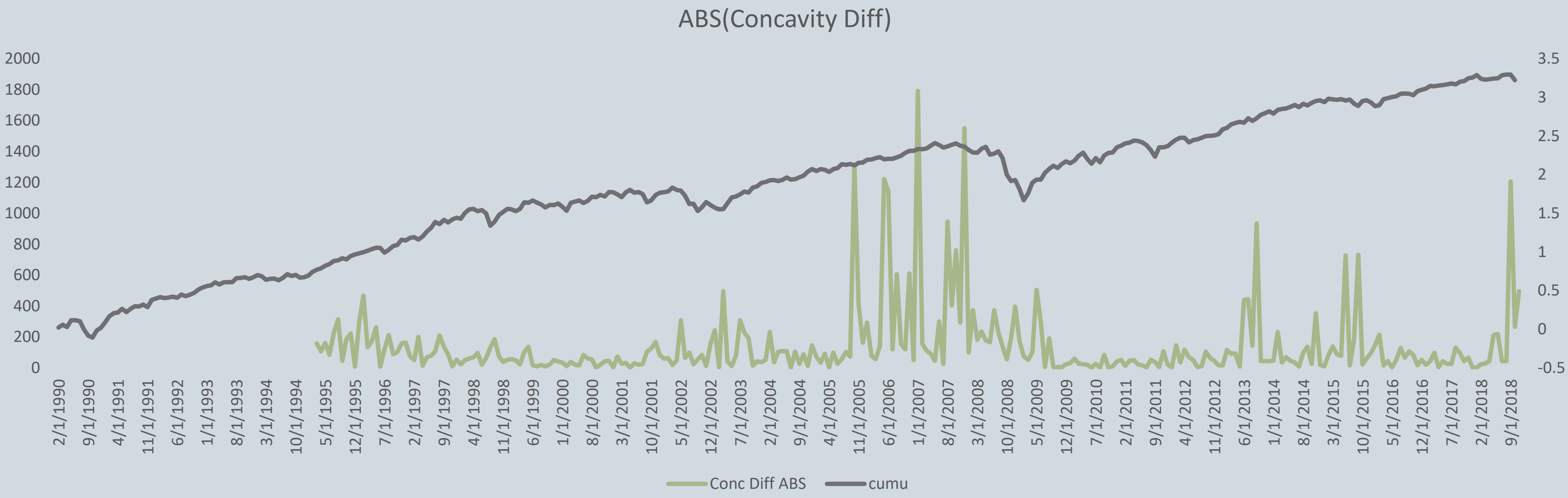
What are favorable and unfavorable conditions to the strategy?

- If not done yet, the strategy should be simulated in the past, through crises and calmer moments, on other markets.
- Do not invest if you don't know when to time in and when to time out.

Monitor the statistical bias that is at the origin of the success

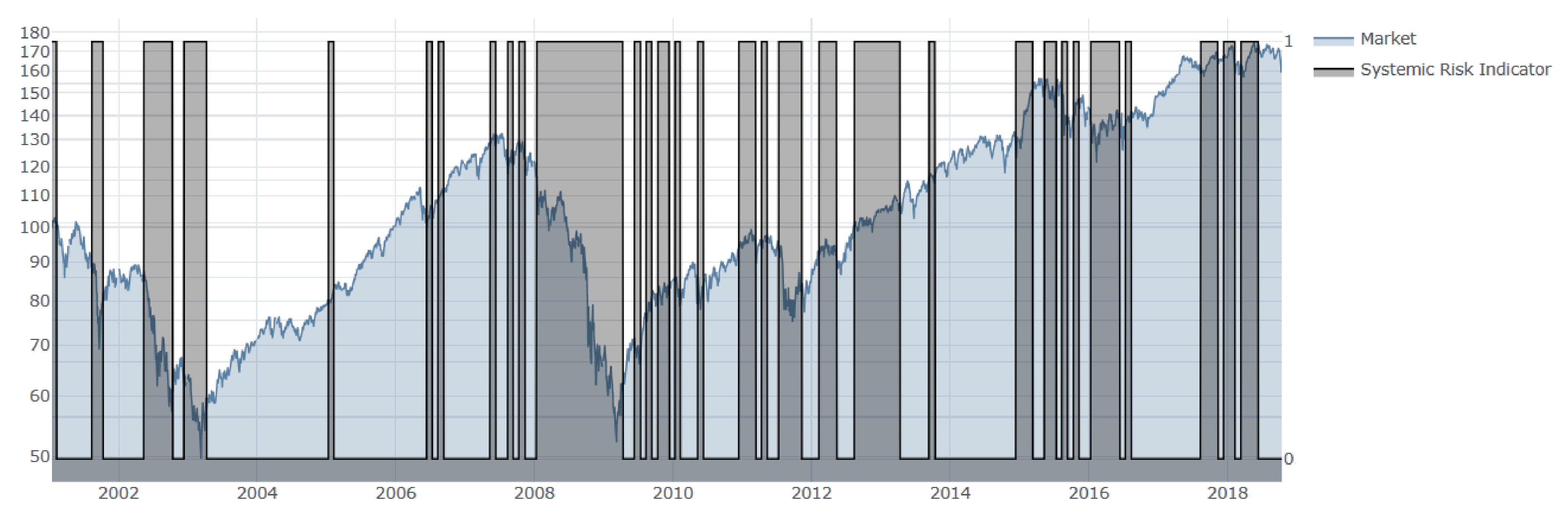
- If it declines, cut the position
- DO NOT set targets, which will force you to reinforce the position if you lose the edge.

Yao Kuang's Crisis Indicator: Nonlinearity



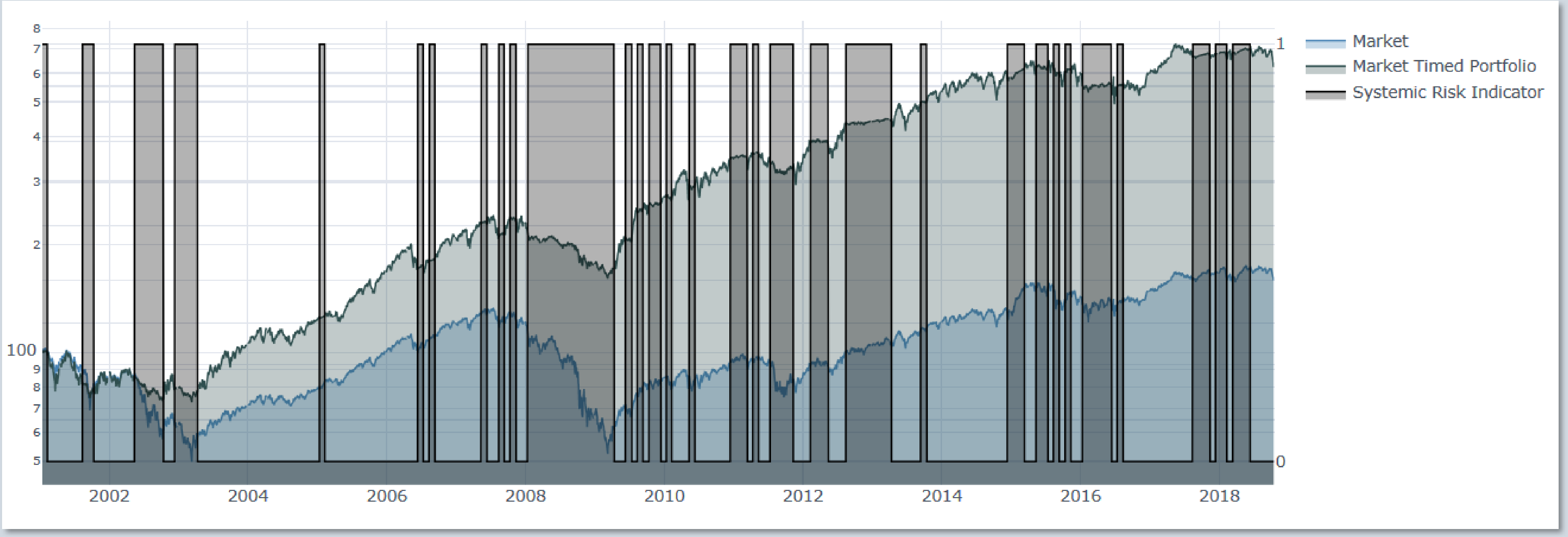
Crisis Indicator = Weighted Sum over Polymodel of |Nonlinear VaR – Linear VaR|

Thomas Barrau's Crisis Indicator



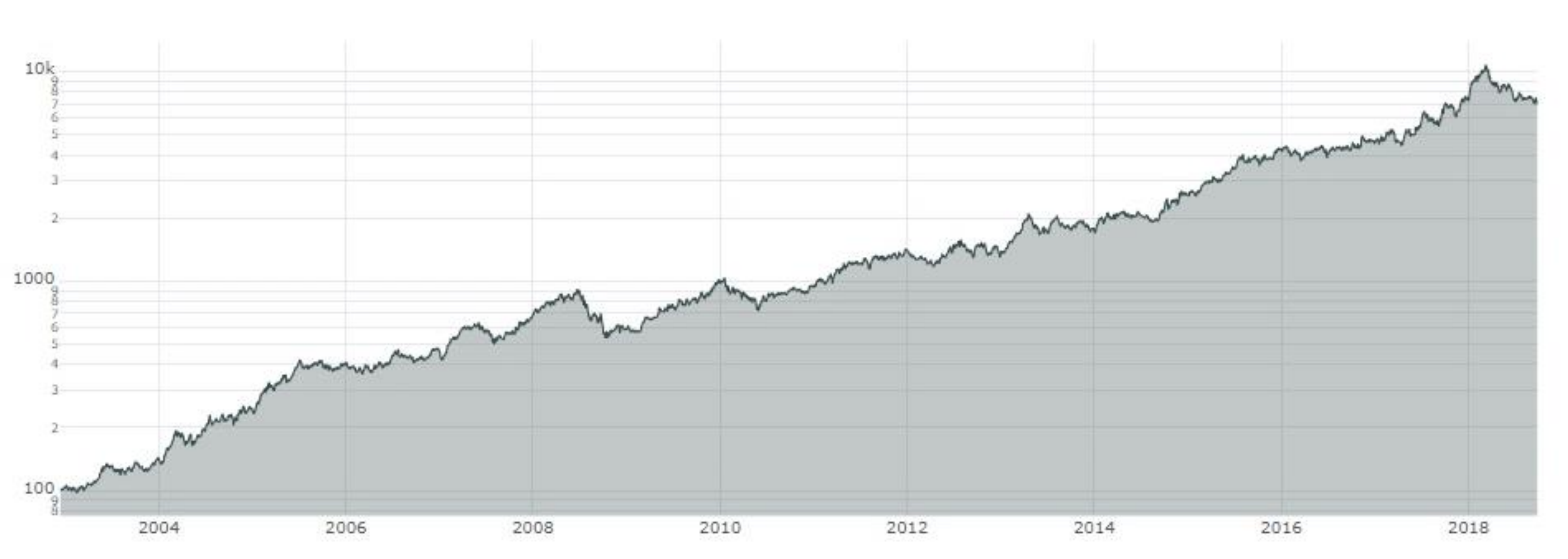
Based on Hellinger distance to pre-crisis distribution

Thomas Barrau's Crisis Indicator



Based on Hellinger distance to pre-crisis distribution

Long-Short Portfolio of Sector Indices

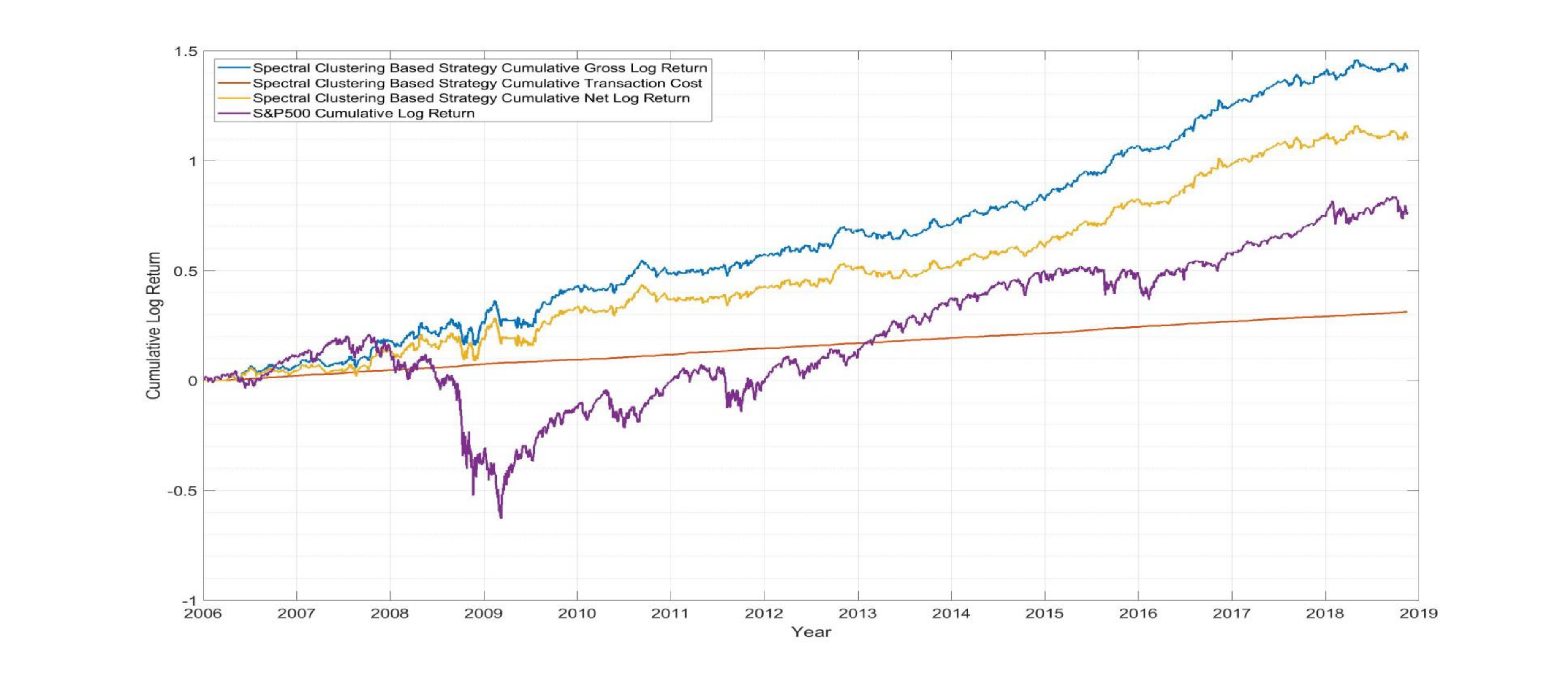


Based on the Antifragility of each sector index

Long-Short Portfolio of Stocks

Based on a cross-sectional prediction of stock returns from their polymodel analysis

Juehui Zhang Pair Trading



Clustering Stocks based on Polymodel Proximity