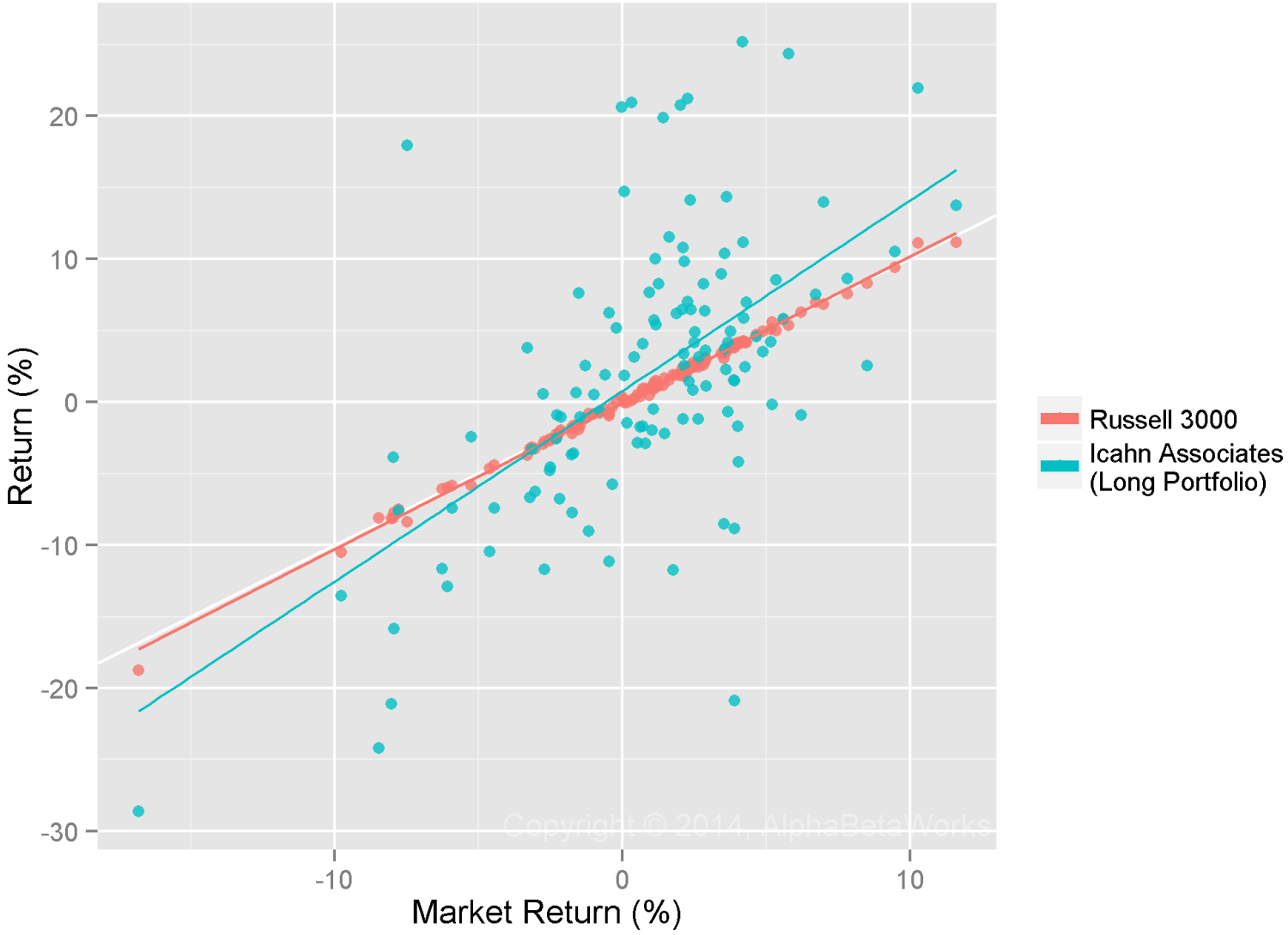
Characterizing Hedge Fund Clustering

Greg Kapoustin, Principal

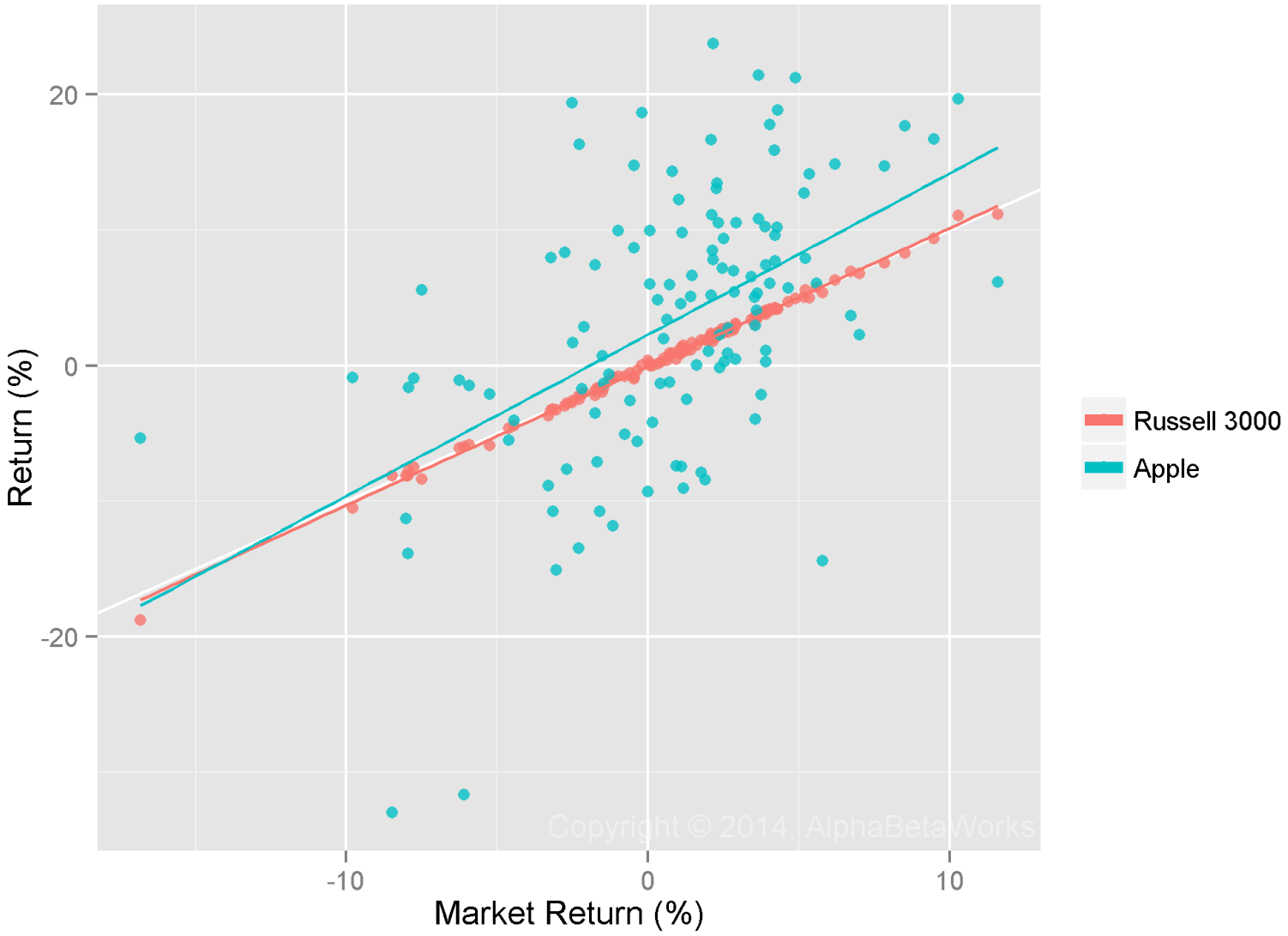
AlphaBetaWorks

gregk@alphabetaworks.com

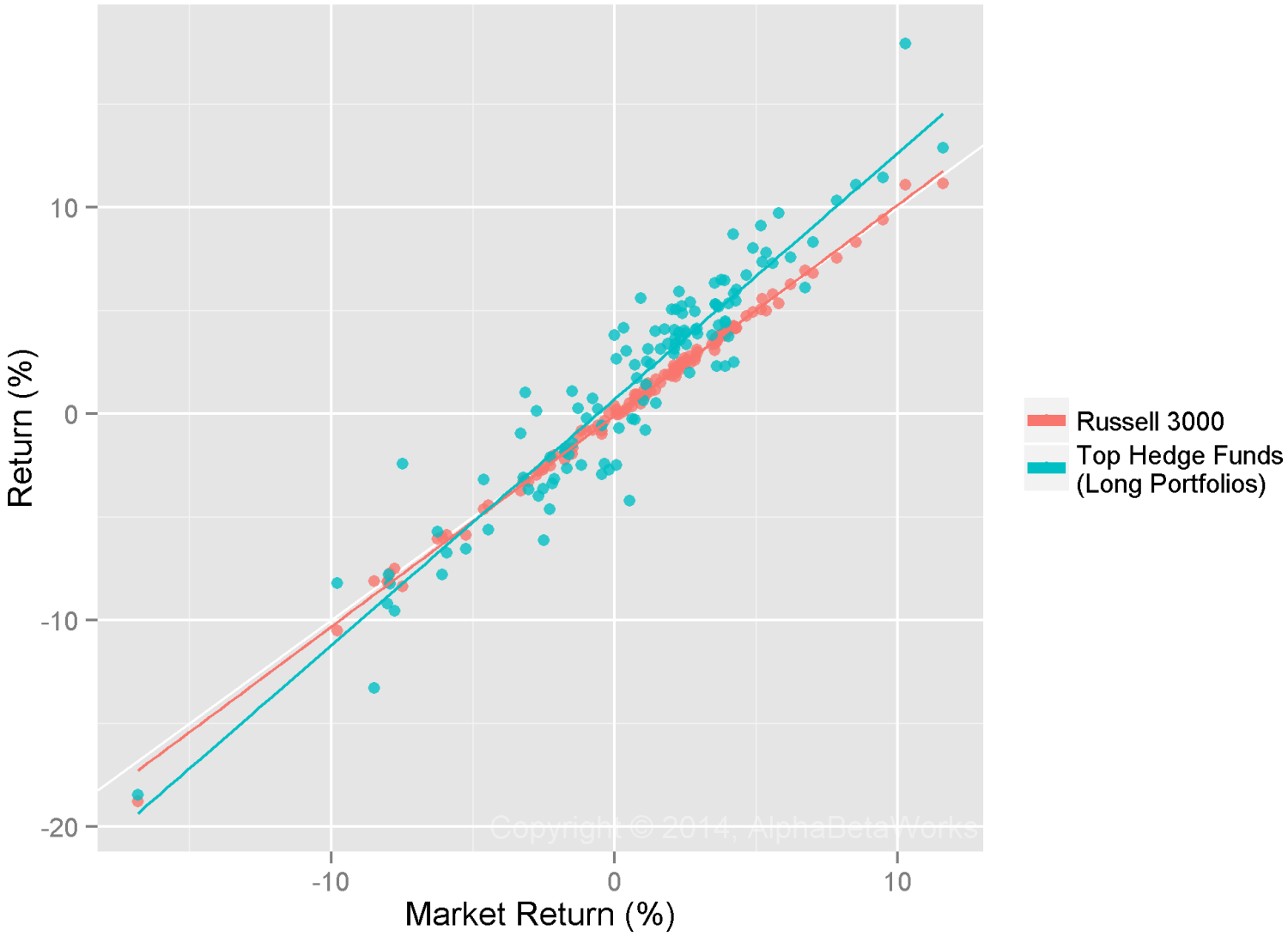
Why do Hedge Funds Exist?



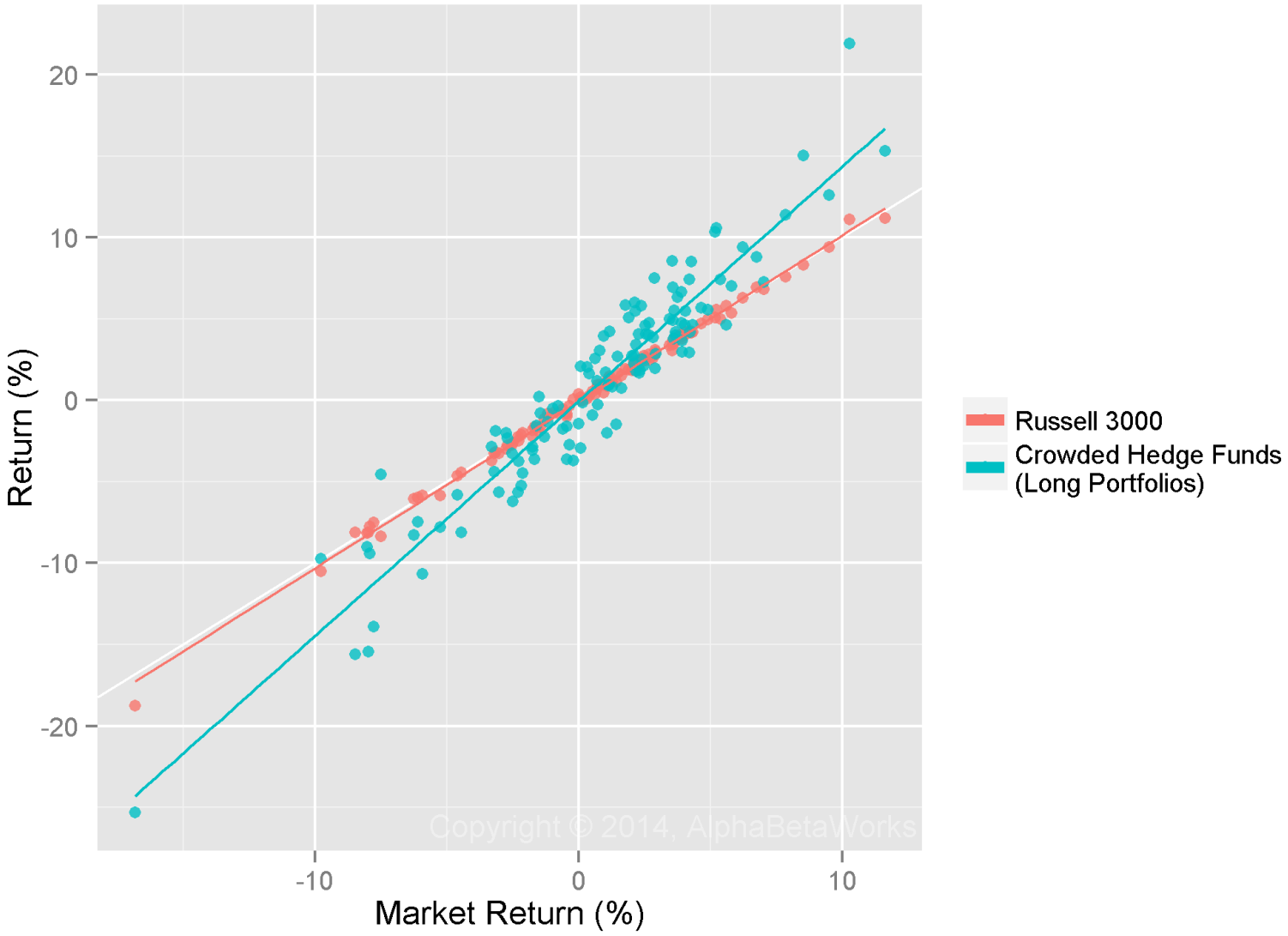
What is Alpha?



What Investors Want



What Investors (Often) Get



Put Differently

What Investors Want

What Investors Get

Volatility as Distance

$$\sigma_p = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \, \sigma_i \sigma_j \rho_{ij}}$$

$\sigma_i \sigma_i \rho_{ii} = \sigma_i^2$ is the variance of asset i

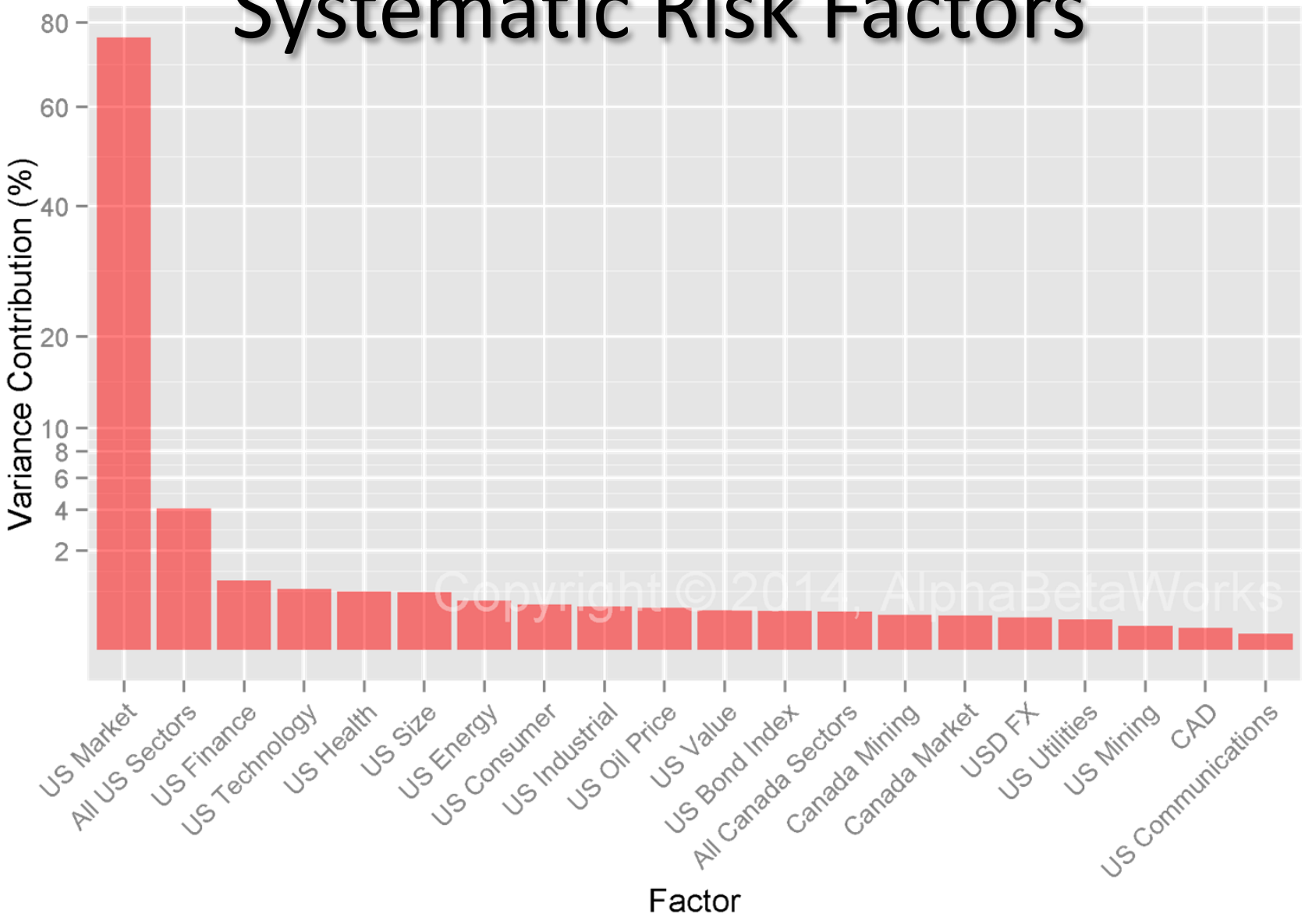
$\sigma_i \sigma_j \rho_{ij} = \sigma(i,j)$ is the covariance between assets i and j.

Factor and Residual Distances

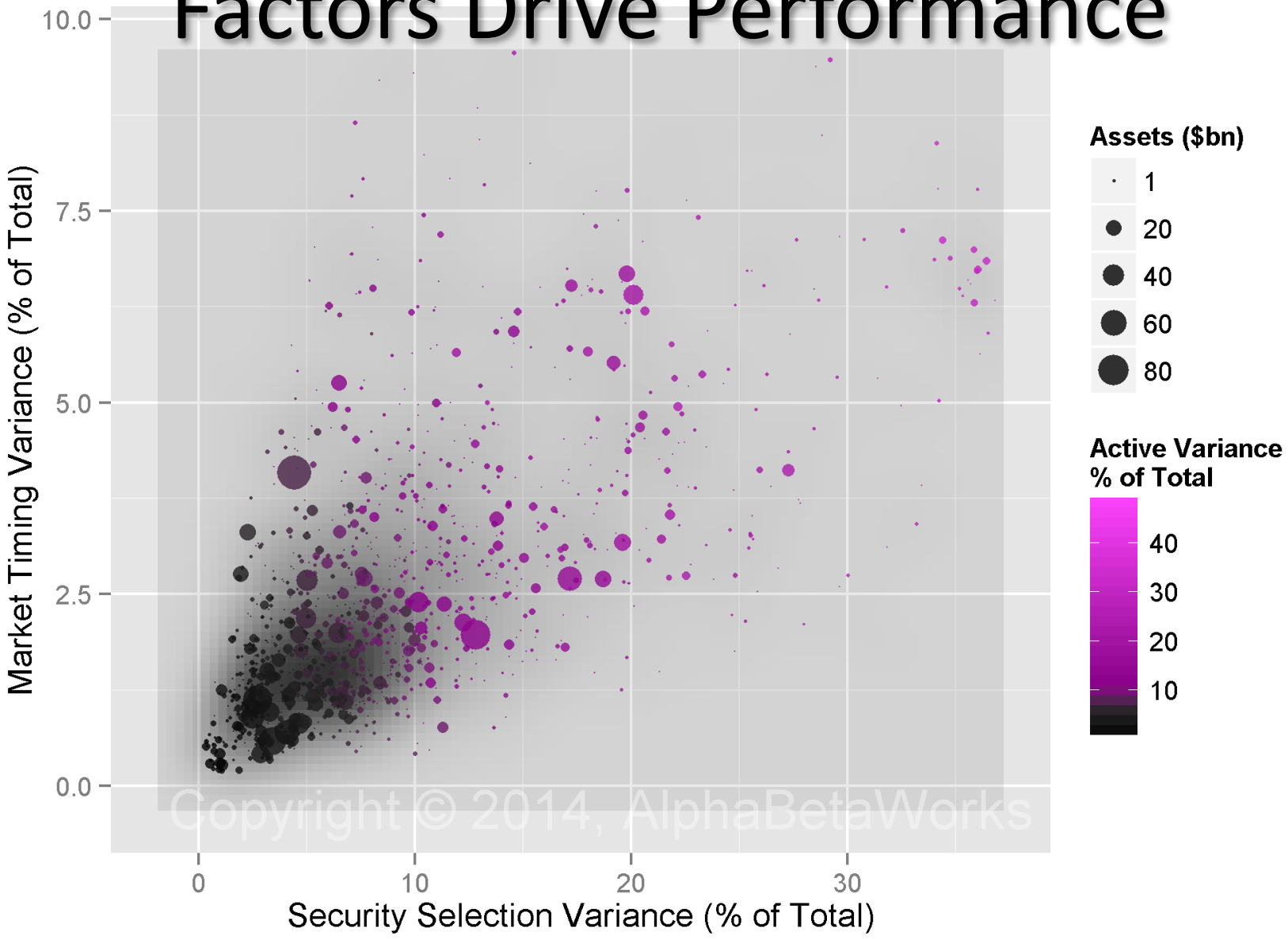
$$\sigma_p = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \, \sigma_i \sigma_j \rho_{ij}}$$

$$= \sqrt{\sum_{i=1}^{n} w_i^2 \sigma_{residual_i}^2 + \sum_{i=1}^{f}\sum_{j=1}^{f} w_i w_j \, \sigma_i \sigma_j \rho_{ij}}$$

Systematic Risk Factors



Factors Drive Performance



Estimating Factor Exposures

```r
# Exponentially weight observations:
# decay constant (0.02) (36 month ~50% decay)
kExpWeights <- exp(-1 * c(1:kDefaultRegPeriods) * .02)

# Estimate exposures using iterated re-weighted least squares (IWLS)
# Default psi = psi.huber
library(MASS)
rlm(
  y ~ x1 + x2 + x3,
  weights = head(kExpWeights, length(y)),
  acc = 0.01
)
```

sMed.X Team optimized the rlm() critical path for AlphaBetaWorks:

- Rewrote QR decomposition in MASS, currently implemented in FORTRAN, called through a C wrapper:

```
c original (dqrdc.f) linpack version dated 08/14/78 .
c g.w. stewart, university of maryland, argonne national lab.
c
c this version dated 22 august 1995
c ross ihaka
c
c bug fixes 29 September 1999 BDR (p > n case, inaccurate ranks)
```

- Added support of a response matrix in a call to rlm(), eliminating the call loop.
- Rewrote the QR algorithm in CUDA-optimized code.
- Achieved >200x performance gain on consumer hardware ($130 NVIDIA GTX 750Ti).

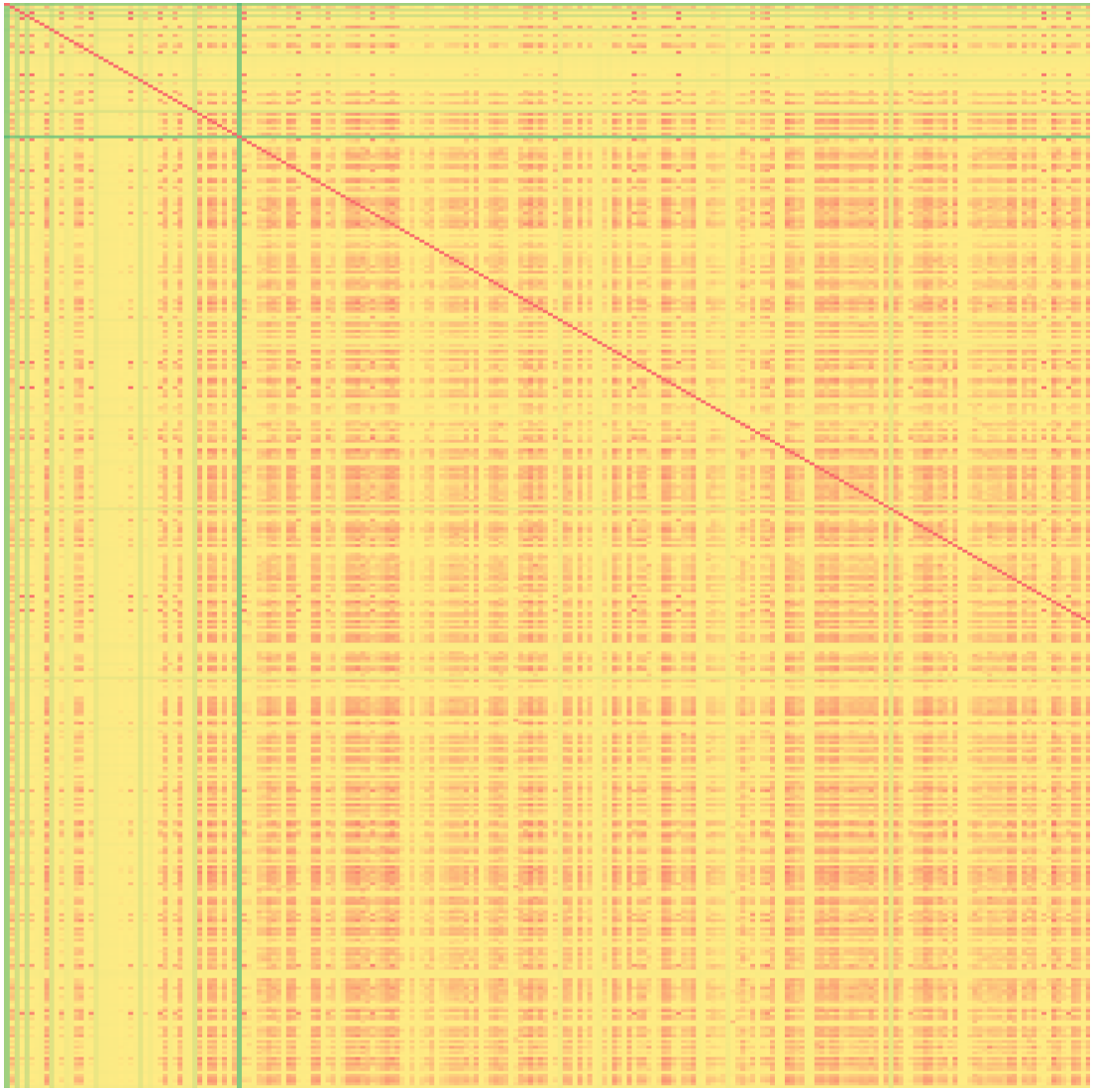
Portfolio Distance Function

```
PortfolioPairTrackingError <- function(port1, port2) {

  # Calculate relative security weightings
  pos1 <- port1$pos.last[which(port1$pos.last != 0)]
  pos2 <- port2$pos.last[which(port2$pos.last != 0)]
  secs1 <- names(pos1)
  secs2 <- names(pos2)
  secs <- union(secs1, secs2)
  port <- setNames(vector("numeric", length(secs)), secs)
  port[secs1] <- pos1[secs1]
  port[secs2] <- port[secs2] - pos2[secs2]

  # Calculate relative residual variance
  var.resid <- sum(port ^ 2 * risk.model$sec.info[secs, "ResidVar"])

  # Calculate relative factor variance
  fact.exps <- port1$fact.exps.last - port2$fact.exps.last
  fact.exp.mat <- fact.exps %*% fact.exps
  var.fact <- sum(fact.exp.mat * risk.model$fact.cov.mat)

  return(sqrt(var.resid + var.fact))
}
```

Calculating Relative Distances

```
library(doParallel)

n.workers <- 36
cl <- makeCluster(n.workers)

registerDoParallel(cl)

tracking.errors <- foreach(
  port.pair = port.pairs,
  .combine = c,
  .errorhandling = 'stop',
  .verbose = TRUE
) %dopar% PortfolioPairTrackingError(
    portfolios[[port.pair$p1]],
    portfolios[[port.pair$p2]]
)

stopCluster(cl)
```

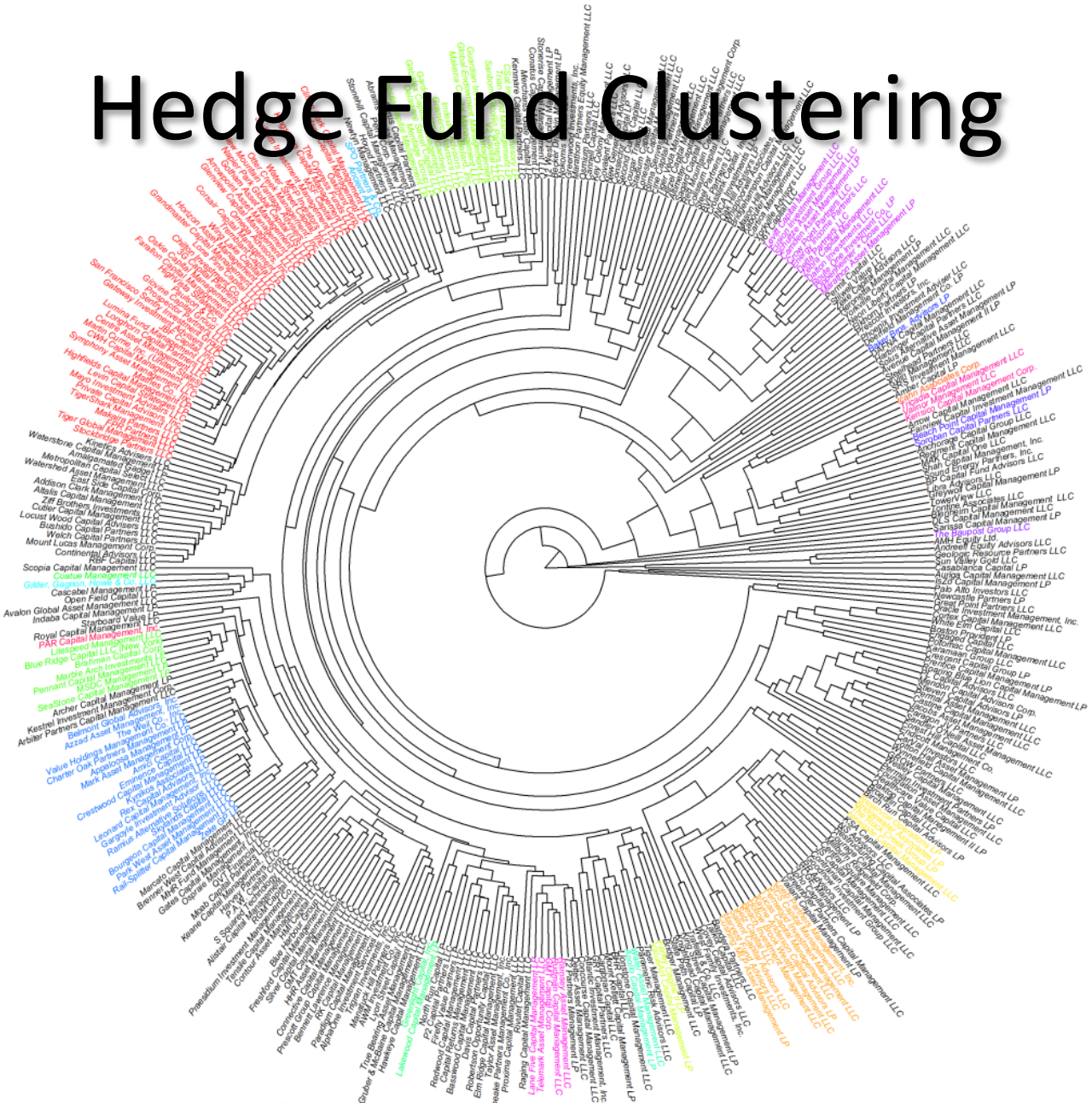
Hedge Fund Distance Matrix



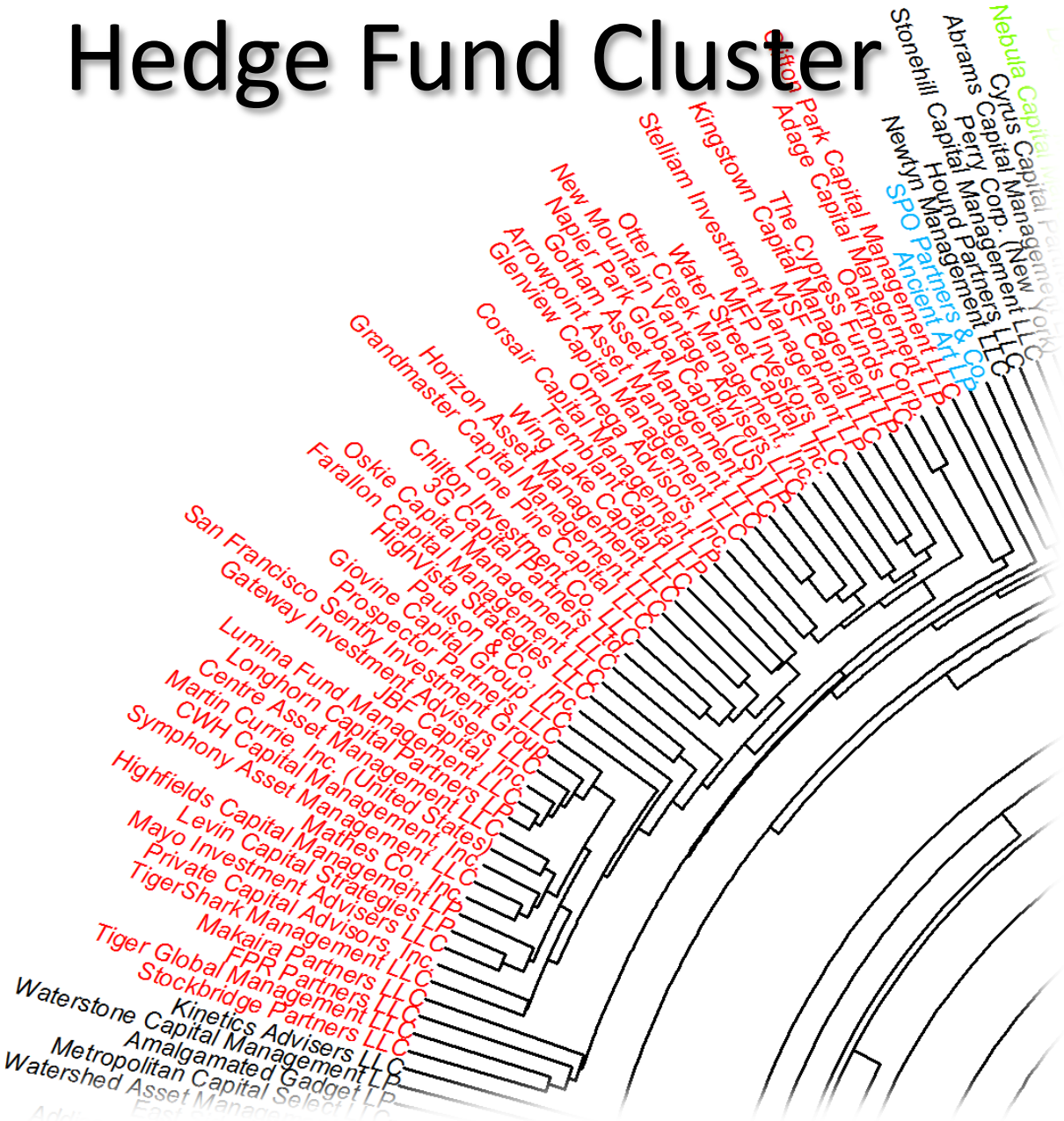
Visualizing the Distance Matrix

```
# Convert tracking error matrix to distance matrix
plot.d.mat <- as.dist(plot.mat)

# Create clusters
hc <- hclust(plot.d.mat, method = "complete")

# Cut dendrogram at h = 10
# Creates clusters with <= 10% forecasted relative tracking errors
clus = cutree(hc, h = 10)

# Plot phylogenic tree of hedge fund portfolios
pal <- rainbow(20)
plot(
  as.phylo(hc),
  type = "fan",
  cex = 0.5,
  tip.color = pal[clus],
  label.offset = 0.01
)
```

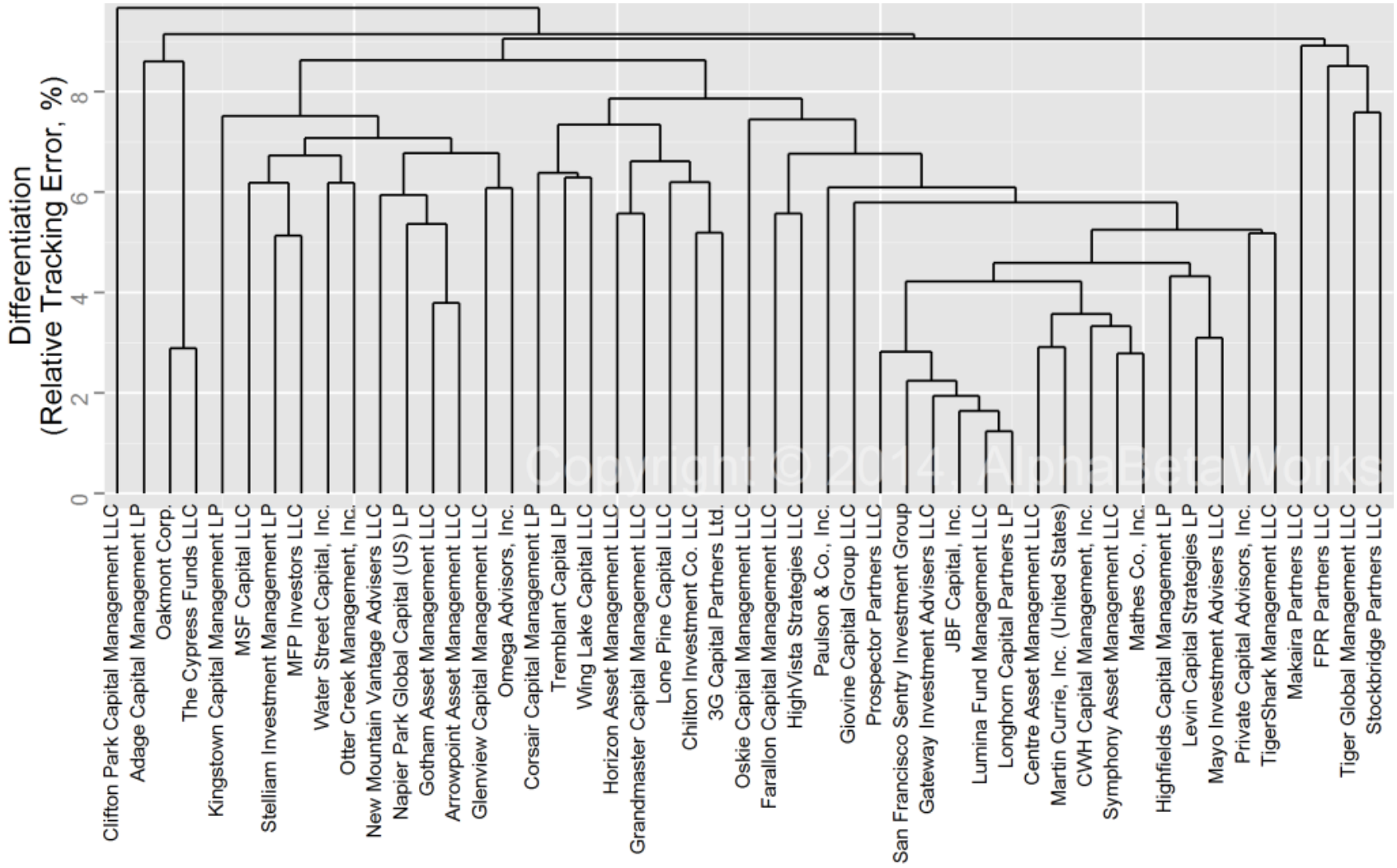
Hedge Fund Clustering



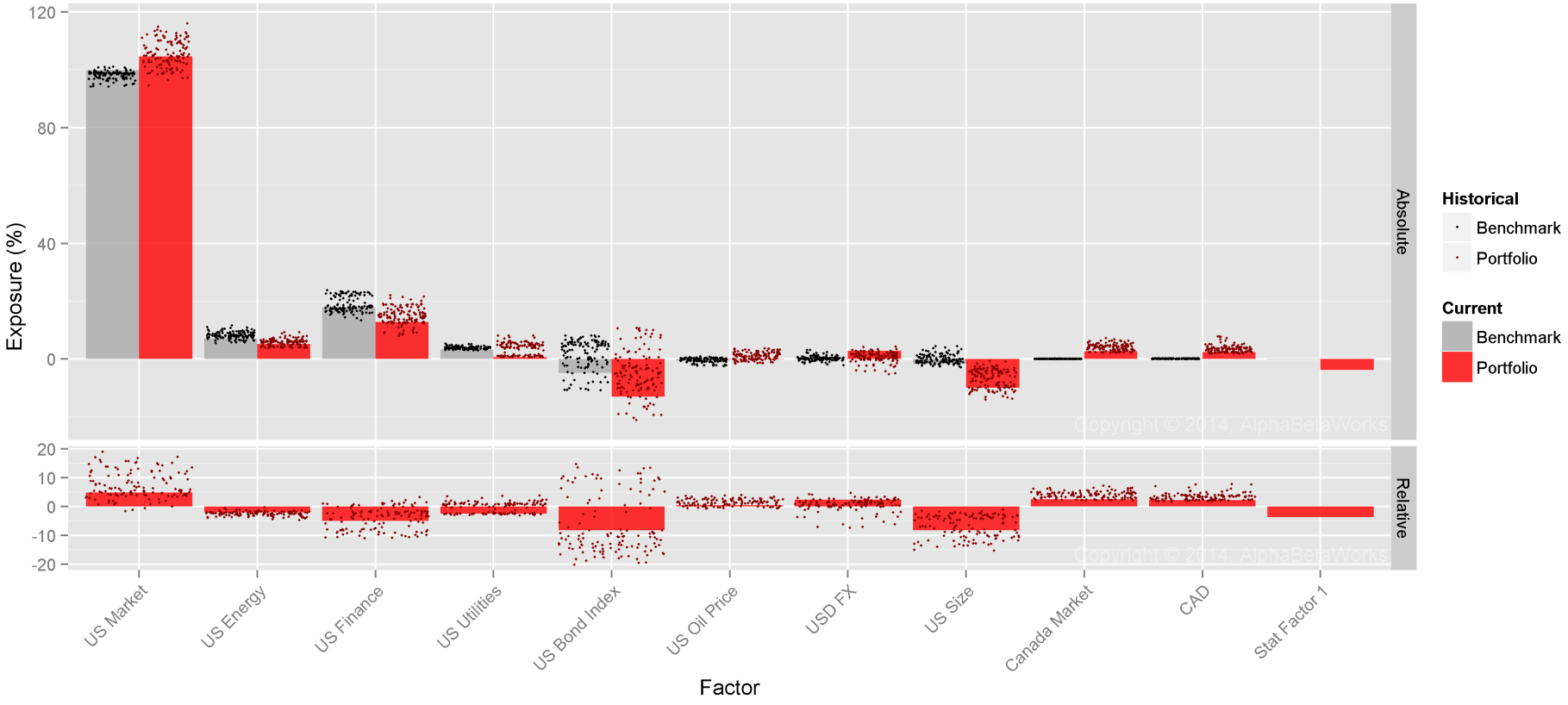
Hedge Fund Cluster



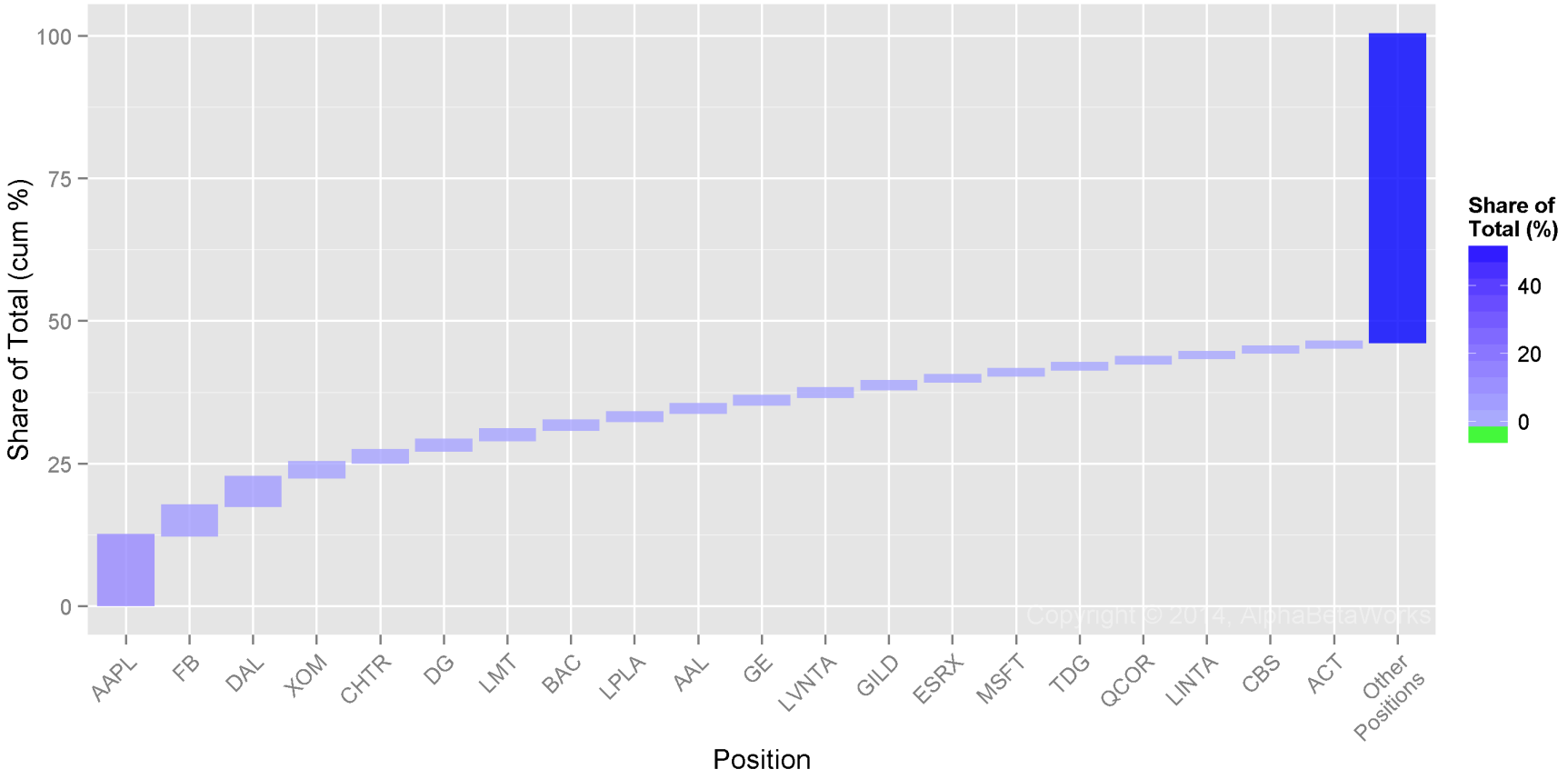
Hedge Fund Cluster



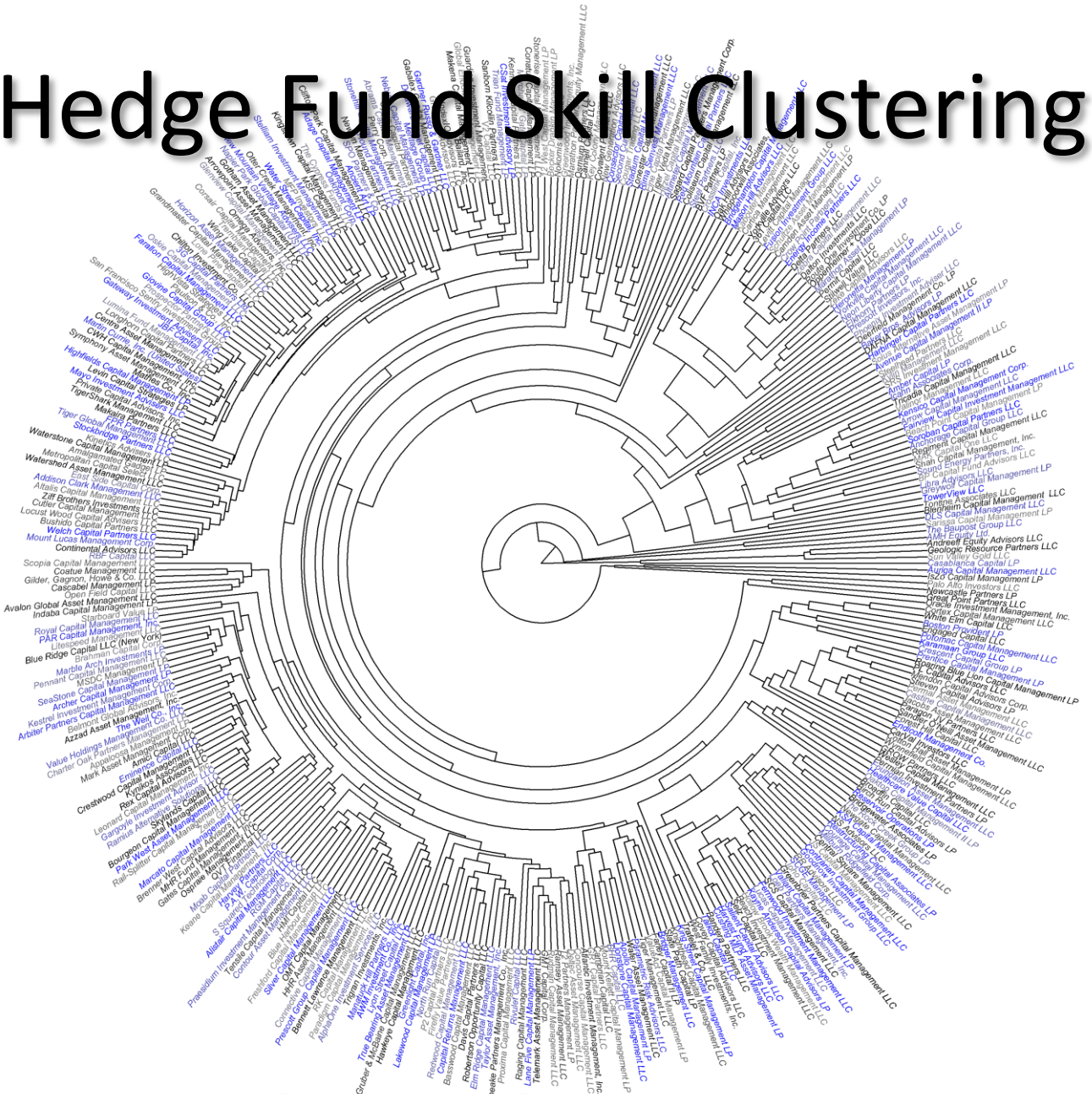
Cluster's Factor Bets



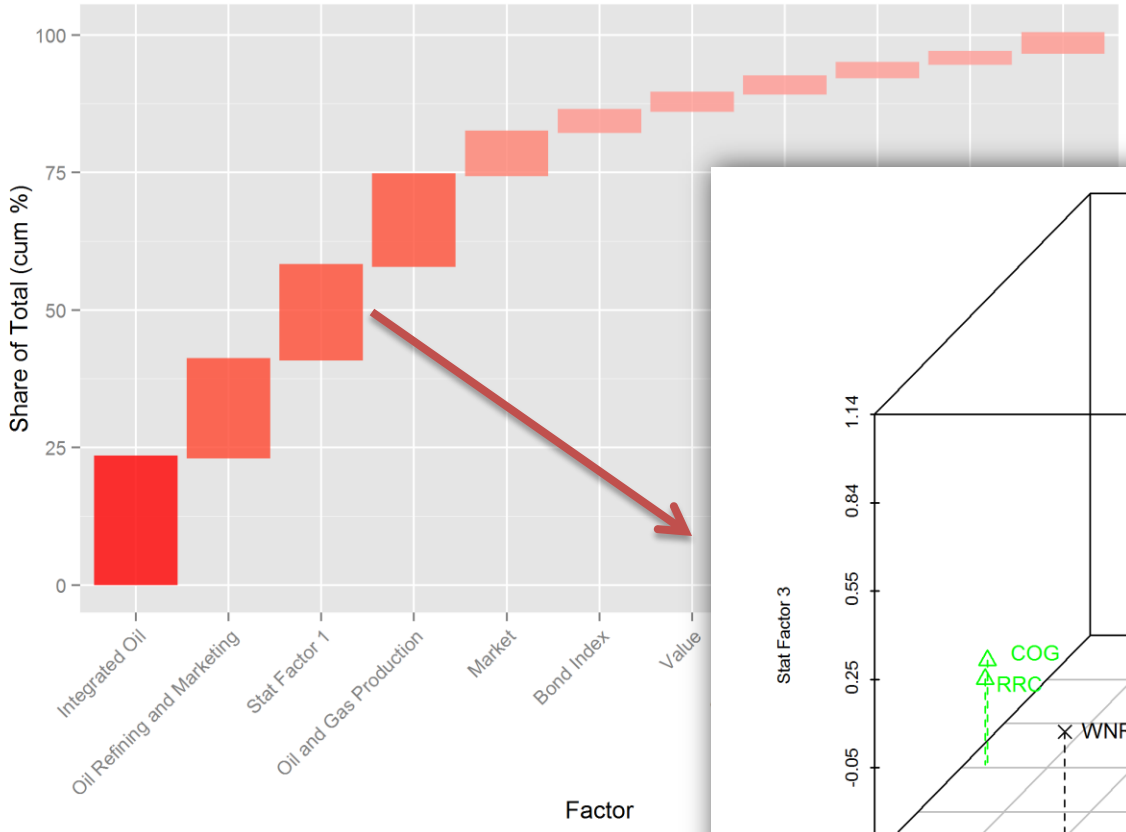
Cluster's Residual Bets

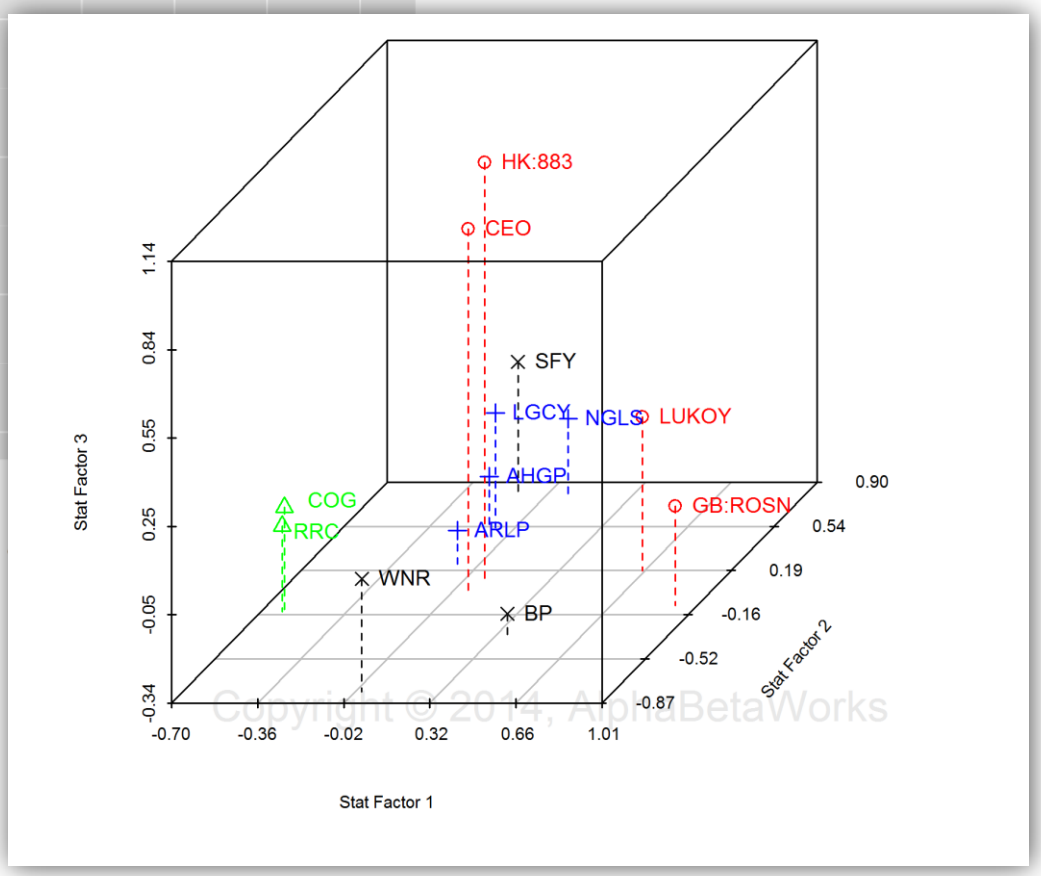


Hedge Fund Skill Clustering

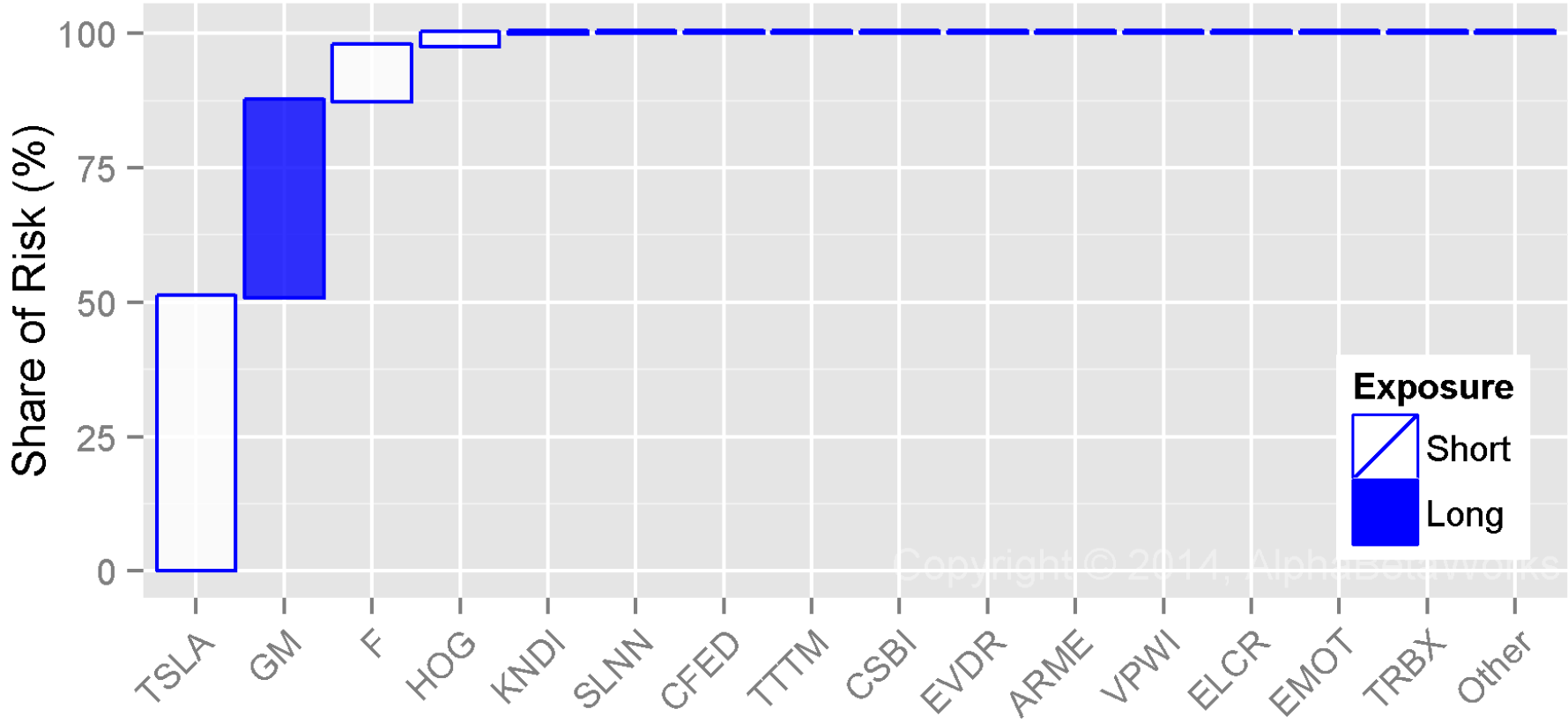


Cluster-Specific Factors





Scary Cluster



Thank You!

Resources

- Wikipedia. *Modern Portfolio Theory.* http://en.wikipedia.org/wiki/Modern_portfolio_theory
- Wikipedia. *Phylogenic Trees.* http://en.wikipedia.org/wiki/Phylogenetic_tree
- Wikipedia. *Cluster Analysis*. http://en.wikipedia.org/wiki/Cluster_analysis
- G. James, D. Witten, T. Hastie, R. Tibshirani (2014). *An Introduction to Statistical Learning.* Springer
- P. J. Huber (1981). *Robust Statistics*. Wiley.
- A. Marazzi (1993). *Algorithms, Routines and S Functions for Robust Statistics*. Wadsworth & Brooks/Cole.
- W. Venables and B. Ripley (2002). *Modern Applied Statistics with S.* Fourth edition. Springer.
- D. Lawley and A. Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method.* Second edition. Butterworths.
- Steve Weston and Rich Calaway (2014, February 26). *Getting Started with doParallel and foreach.* http://cran.r-project.org/web/packages/doParallel/vignettes/gettingstartedParallel.pdf
- Steve Weston (2014, April 10). *Using the foreach Package.* http://cran.r-project.org/web/packages/foreach/vignettes/foreach.pdf
- AlphaBetaWorks (2014, October 20). *Hedge Fund Clustering.* http://abwinsights.com/2014/10/20/hedge-fund-clustering/
- AlphaBetaWorks Insights and AlphaBetaWorks Charts. *Applied Financial Analytics.* http://abwinsights.com http://abwcharts.com

Disclaimer

This material has been prepared by Alpha Beta Analytics, LLC, d/b/a AlphaBetaWorks. This material is for information and illustrative purposes only. It is not, and should not be regarded as investment advice or a recommendation regarding any particular investment, fund, or course of action. Opinions expressed herein are current opinions as of the date appearing in this material only and are subject to change without notice. No representation or warranty is being made that the information herein is accurate, correct, complete, or timely. This information is provided with the understanding that with respect to the material provided herein, you will make independent decisions with respect to any course of action in connection herewith. By accepting this material, you acknowledge, understand and accept the foregoing:

THE PROJECTIONS, FINANCIAL OR OTHERWISE, SET FORTH HEREIN, ARE ESTIMATIONS BASED ON CURRENT INFORMATION AND RELY ON A NUMBER OF ASSUMPTIONS CONCERNING FUTURE EVENTS AND ARE SUBJECT TO A NUMBER OF UNCERTAINTIES AND OTHER FACTORS, MANY OF WHICH ARE OUTSIDE THE CONTROL OF ALPHABETAWORKS. SUCH FACTORS COULD CAUSE ACTUAL RESULTS TO DIFFER MATERIALLY FROM WHAT IS SET FORTH IN THIS MATERIAL. ALPHABETAWORKS UNDERTAKES NO OBLIGATIONS TO UPDATE OR REVISE ANY OF THE PROJECTIONS, WHETHER AS A RESULT OF NEW INFORMATION, FUTURE EVENTS OR OTHERWISE. ALPHABETAWORKS HEREBY DISCLAIMS ANY EXPRESS OR IMPLIED REPRESENTATIONS OR WARRANTIES, WHETHER AT LAW OR IN EQUITY, INCLUDING AS TO THE ACCURACY OR COMPLETENESS OF, OR REASONABLENESS OF ANY ASSUMPTIONS UNDERLYING ANY ESTIMATES, PROJECTIONS AND FORECASTS SET FORTH, ANY INFORMATION, DOCUMENTS, OR MATERIALS REGARDING ALPHABETAWORKS (FURNISHED OR MADE AVAILABLE IN THIS MATERIAL OR OTHER MATERIALS PROVIDED BY ALPHABETAWORKS).