Performance Analysis in R

R/Finance Workshop

Peter Carl
Brian Peterson
Authors of *PerformanceAnalytics*

April 24, 2009
Chicago, IL

- ## Introduction

  - About PerformanceAnalytics

- ## Performance Analysis

  - Types of Analysis

- ## Measures and Calculations

  - Measuring Return

  - Assessing Risk

  - Evaluating Relative Performance

  - Examining Style Exposures

- ## Appendix: Building Blocks

  - Data

  - Graphics

  - Tables

- Library of econometric functions for performance and risk analysis of financial portfolios.

- Aims to be useful to both practitioners and researchers alike.

- Analysis of return streams, whether distributed normally or not.

- In development since early 2006, packaged and first released to CRAN in 2007.

- Now contains more than 180 functions and more than 11,000 lines of code and 7,500 lines of documentation.

- Collaboration, patches and suggestions from users in industry and academia worldwide

CRAN Version 0.9.7.1

- Sharpe's Style Analysis

- Snailtrail chart

- VaR Sensitivity chart

- Modified Expected Shortfall

- Multivariate moments and risk metrics

- Higher co-moments

- Robust data cleaning

- Many fewer dependencies

- Other new functions, bug fixes

Release 1.0

- Component VaR and

- Component Expected Shortfall

- Wrappers of VaR and ES functions for consistency

- Adoption of xts for time series

- Improved x-axis handling in charts

- Formatted tables in devices

- Capture ratios and other metrics
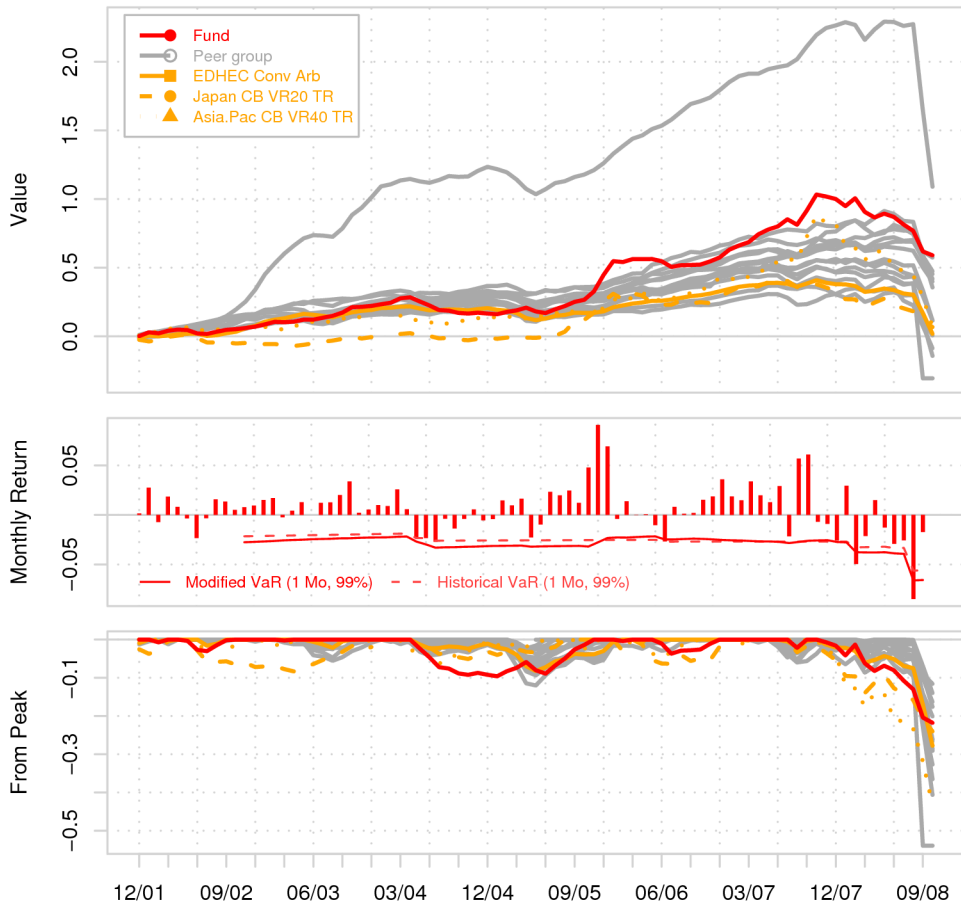
- Bug fixes, new functions

What

- What are we trying to accomplish?

- Accrue evidence to help ask better questions

  - Is this something I want in my portfolio? What is the risk of investing? How do I maximize my expected gain and minimize my chances of going broke? Is this manager worth the fees he charges? Is there a better substitution? What was happening when...?

- Measurement, not prediction

- Consider return and risk together

- Handle small, biased samples of expensive data

- Comparisons can be tenuous, be careful of framing

- Only one of a set of tasks taken to understand a current or potential investment
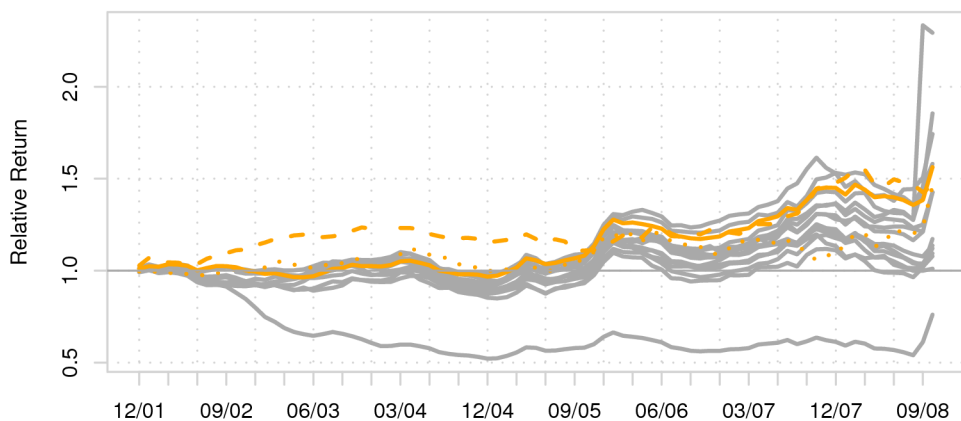
- Shows a set of hedge funds within a particular strategy

- Is typical of situations faced by investors

  - Small, biased sample of live, open funds

  - Peer group have return series of differing lengths and quality

  - Too small to make confident statistical observations about

  - Too large to present easily

  - Returns are clearly non-normal, non-stationary

  - Focus is on a single manager with the context provided by peers and a few indexes

- Given that the assumptions are strained, what can we observe about performance and risk?

- Caveats: backwards looking, no silver bullet, many "hidden" risks, your mileage may vary, more than one way to skin a cat, etc.

Fund Performance

Legend:
- Fund
- Peer group
- EDHEC Conv Arb
- Japan CB VR20 TR
- Asia.Pac CB VR40 TR

Modified VaR (1 Mo, 99%) — — — Historical VaR (1 Mo, 99%)
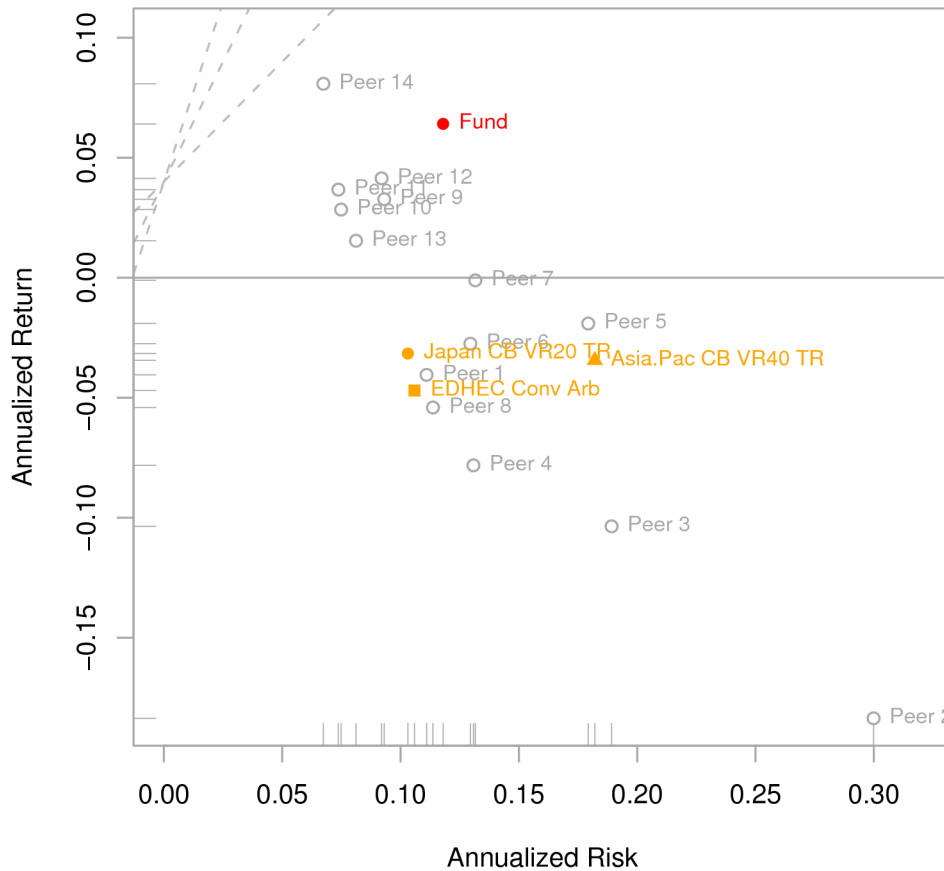
Relative Performance

- How has the fund/ strategy/ portfolio performed in the past?

- Cumulative performance does not show relative performance well, but gives a sense for the overall shape. See **chart.CumReturns**

- Ranking and relative performance along the way is impossible to assess.

- Drawdowns show extent of losses relative to peak 'equity'. See **chart.Drawdown**

- VaR through time shows how estimation changed. See **chart.BarVaR**

- Relative performance shows periods of out-performance, under-performance. See **chart.RelativePerformance**
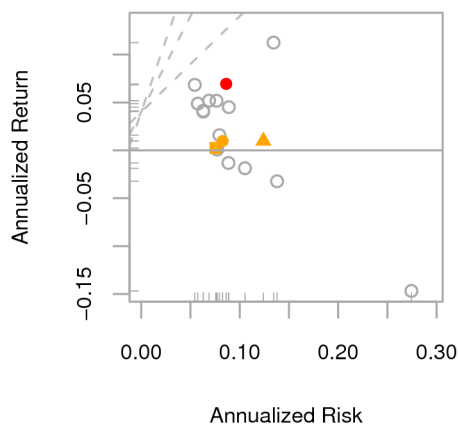
Performance During Trailing 36–Months

Since Inception of Fund

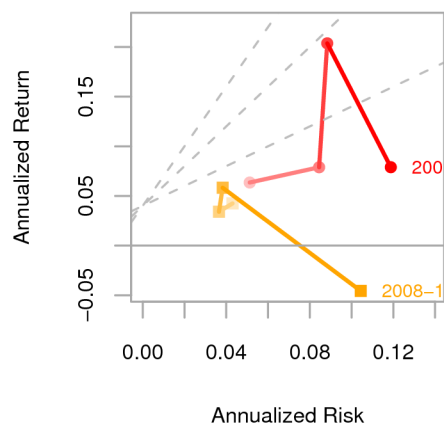**Trailing 36–Months Calc'd 12 Months**

- How do return and risk compare?

- Traditional view normalizes return and volatility by annualizing the measures. See **chart.RiskReturnScatter**

- Short histories are a problem, and "from inception" view masks differing time periods

- Trailing period performance provides some smoothing

- Dotted lines show Sharpe ratio

- Snail trail shows time series overlaid on return and risk. See **chart.SnailTrail**

- How has the fund's relative performance and risk position changed through time?

36-Month Rolling Performance
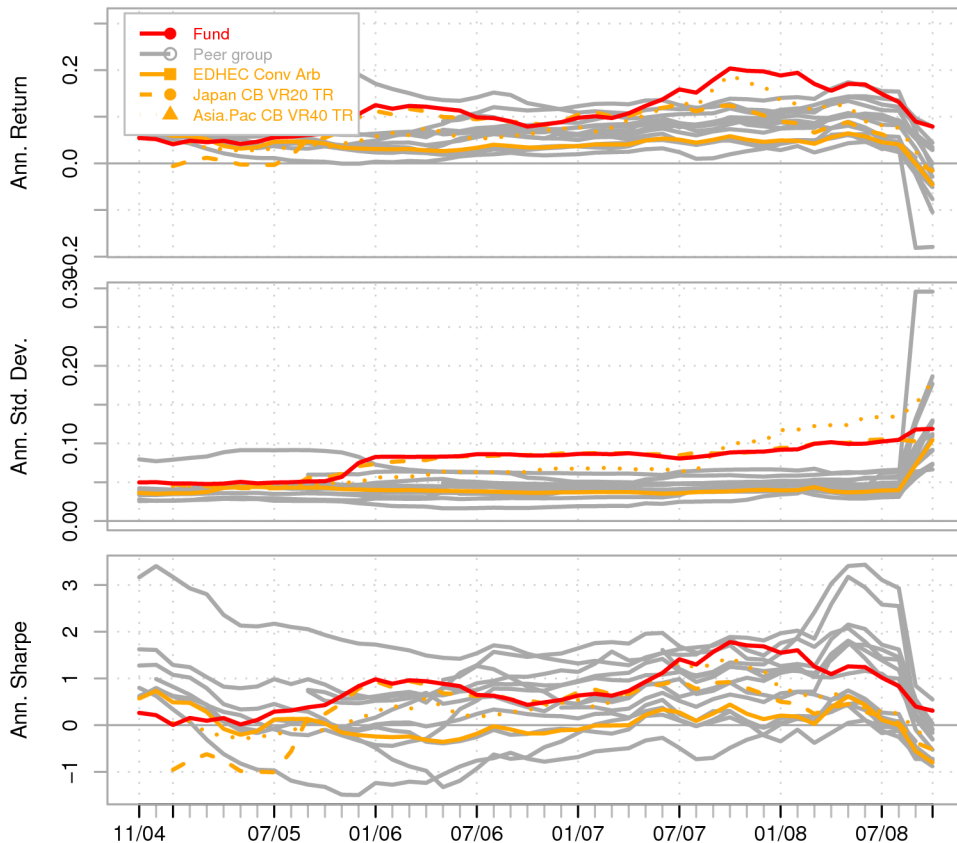
Legend:
- Fund
- Peer group
- EDHEC Conv Arb
- Japan CB VR20 TR
- Asia.Pac CB VR40 TR

Rolling 36-Month Correlation to Fund
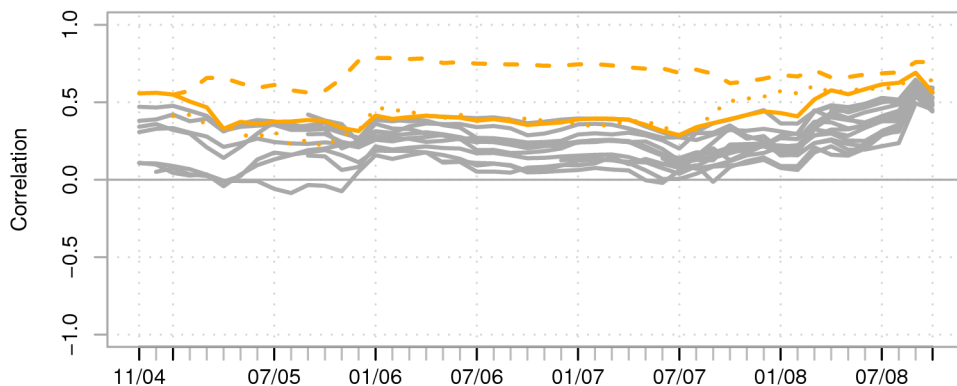
- · How consistent has performance been through time?

- · Has the manager taken more risk to get higher returns through time?

- · Rolling windows provide another view of performance and risk through time, one that is easier to associate with events

- · Reasonable for longer-term assessment

- · Smoothing can help reduce noise, but may also hide short-term issues

- · Use multiple time-frames, consider your investment horizon

- · See **chart.RollingPerformance, chart.RollingCorrelation**

Rolling 36−Month Regression versus EDHEC Conv Arb

Benchmarking and Regression
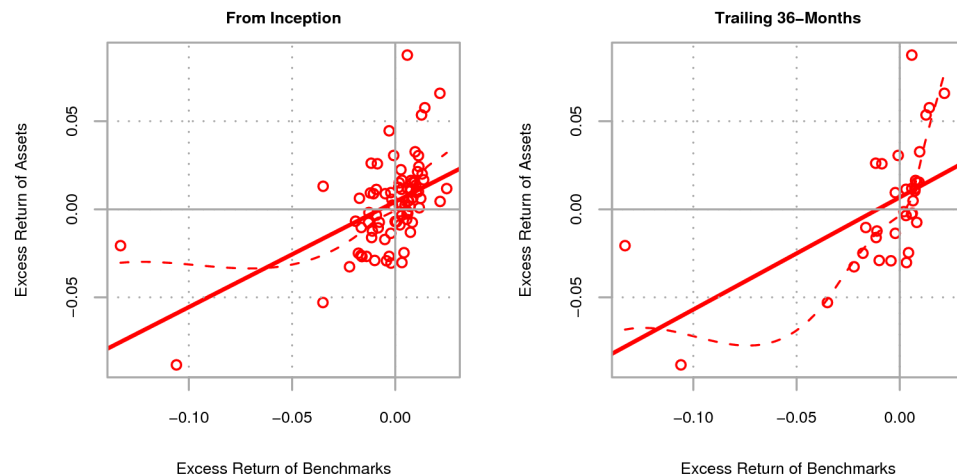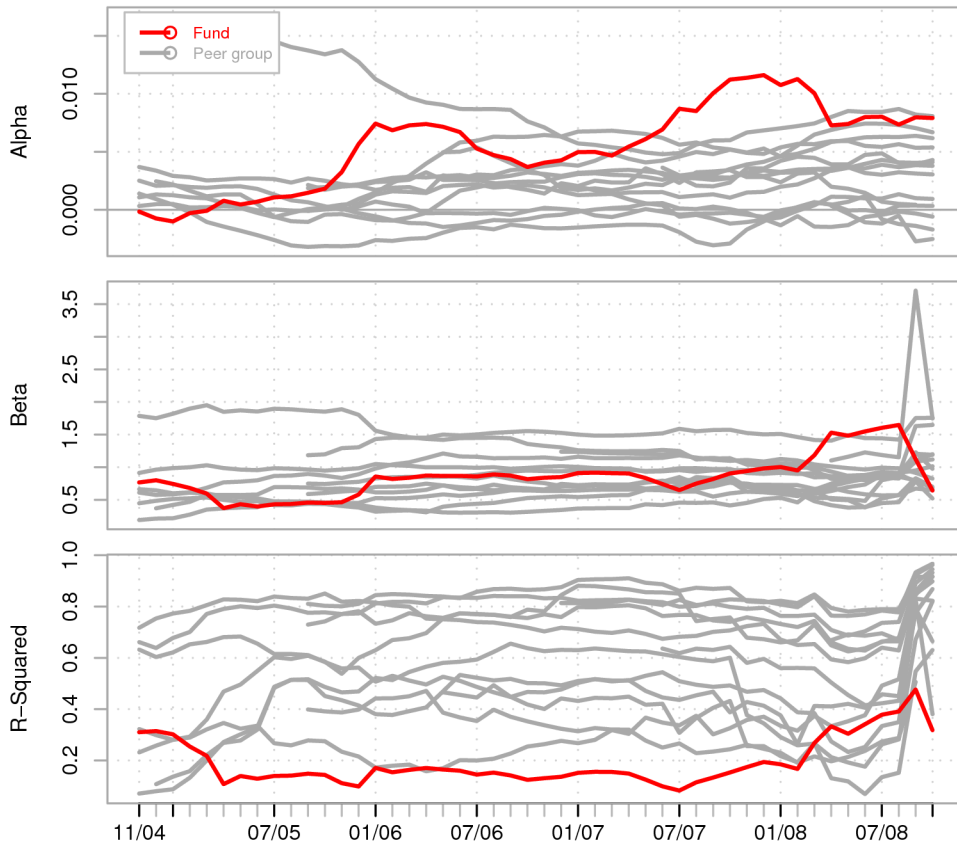
- What benchmark provides a good comparison?

- How has the strategy done against the benchmark? Would we prefer to hold the benchmark?

- How has exposure to the factor changed through time?

- Simple linear regression over rolling periods provides a view of parameter stability, with caveats from prior slide

- Loess fit shows deviations from linear fit

- See **chart.Regression, chart.RollingRegression**
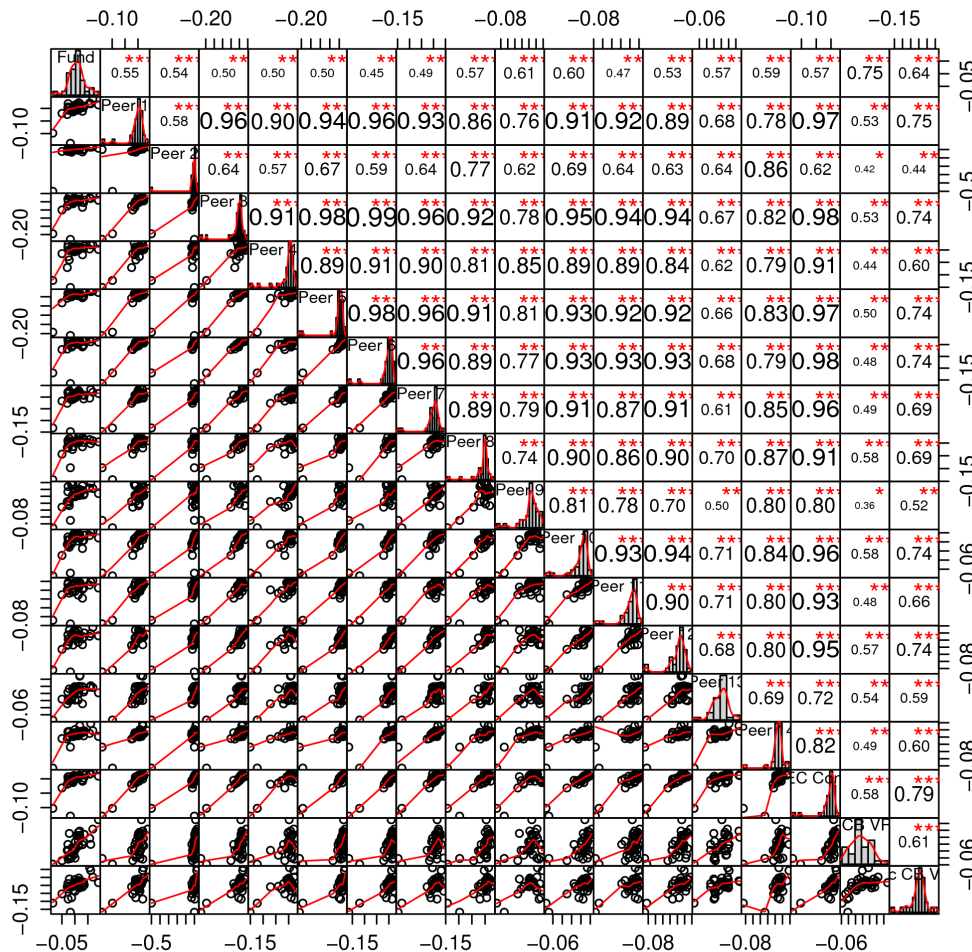
- A high correlation value may be the result of chance and not significant

- A low correlation value may be significant, a small but still meaningful relationship

- Is diversification available? Covariance and correlation are critical concepts in portfolio construction

- How stationary is correlation through time?

- See **chart.Correlation, table.Correlation, chart.RollingCorrelation**

Fund

ACF / PACF plots (Lag)

	rho1	rho2	rho3	rho4	rho5	rho6	Q(6) p–value
Fund	0.4125	0.1964	0.1812	0.1166	0.1288	−0.0691	5e−04
Peer 1	0.4510	0.0668	0.1496	0.1062	0.0204	0.0431	0.0018
Peer 2	−0.0134	0.0390	0.0640	2e−04	−0.0141	0.0550	0.9985
Peer 3	0.4696	0.0274	0.0652	−0.0219	−0.1525	0.0185	0.0040
Peer 4	0.5328	0.2637	0.1901	−0.0197	−0.0389	0.0185	0.0000
Peer 5	0.5611	0.1679	0.1565	0.0791	0.0015	0.0790	0.0000
Peer 6	0.4750	0.0473	0.1136	0.0553	−0.0557	0.0403	0.0185
Peer 7	0.4437	0.1136	0.1809	0.0399	−0.1028	0.0202	0.0012
Peer 8	0.3855	0.0800	0.0803	−0.0588	−0.1272	0.1160	0.0096
Peer 9	0.4638	0.0964	−0.0298	−0.0098	0.0100	0.0988	0.0134
Peer 10	0.5669	0.1476	0.1250	0.0523	−0.0071	0.1847	0.0000
Peer 11	0.4990	0.0460	−0.0207	−0.0534	−0.0382	0.0688	0.0255
Peer 12	0.4375	0.1622	0.0807	0.0543	−0.0726	0.0636	0.0023
Peer 13	0.1287	0.0696	0.0581	−0.0418	0.1348	−0.0611	0.7096
Peer 14	0.4375	0.2194	0.1131	−0.0084	−0.0820	−0.0073	0.0011
EDHEC Conv Arb	0.4973	0.1175	0.1380	0.0692	−0.0354	0.0771	3e−04
Japan CB VR20 TR	0.3207	0.1677	0.2202	0.1162	−0.0584	0.0215	0.0103
Asia.Pac CB VR40 TR	0.3138	0.1012	0.0862	0.0857	0.0042	0.0064	0.1075

- What is the liquidity of the portfolio? What influence does the manager have when marking the book?

- Getmansky, Lo, et. al. (2004) offer that autocorrelation may be observed where liquidity is low

- Geltner (1991): also likely with real estate, venture capital, private equity, art, and other difficult to value assets

- Economic interpretation is that portfolio prices are slow to reflect market values

- See **chart.ACFplus, table.Autocorrelation**

- Should we 'adjust' for auto-correlation? What about negative values?

- What have "normal" losses been in a down period? At what level should we set alarms for re-evaluation of an asset in a portfolio?

- Do losses behave differently than gains?

- Markowitz offers semivariance

- Sortino suggests that risk may include shortfall from a goal

	Semi Deviation	Gain Deviation	Loss Deviation	Downside Deviation (MAR=10%)	Downside Deviation (rf=4%)	Downside Deviation (0%)	Maximum Drawdown	VaR (99%)	Beyond VaR	Modified VaR (99%)
Japan CB VR20 TR	2.3	1.5	1.4	2.6	2.4	2.2	−24.0	5.2	5.2	5.2
Peer 10	2.4	0.6	2.3	2.3	2.5	2.9	−16.3	3.5	3.5	5.7
Peer 13	2.0	1.1	1.5	2.1	2.1	2.1	−14.7	3.9	3.9	6.1
Peer 11	2.6	0.6	2.5	2.5	2.6	3.0	−14.0	3.9	3.9	6.2
Fund	2.5	1.8	1.7	2.6	2.4	2.4	−21.8	5.2	5.2	6.6
Peer 14	2.0	0.8	2.1	1.9	2.2	2.7	−11.6	3.1	3.1	6.8
Peer 12	2.5	0.8	2.3	2.6	2.7	2.8	−17.6	4.1	4.1	7.3
Peer 9	2.8	1.0	2.3	2.7	2.8	3.0	−20.1	4.7	4.7	7.6
EDHEC Conv Arb	2.9	0.7	2.7	2.6	2.9	3.1	−27.8	3.9	3.9	8.5
Peer 8	2.8	0.7	2.8	2.5	2.9	3.2	−26.6	3.9	3.9	9.1
Peer 4	4.2	0.6	3.7	3.3	3.7	4.3	−32.7	5.9	5.9	9.1
Peer 1	2.7	0.9	2.7	2.5	2.7	3.2	−26.1	4.2	4.2	10.1
Peer 6	4.6	0.6	4.1	4.1	4.5	4.7	−28.6	7.2	7.2	10.9
Peer 7	3.4	1.6	3.8	3.5	3.8	4.3	−29.1	5.5	5.5	12.5
Asia.Pac CB VR40 TR	3.8	2.2	3.1	3.9	3.8	3.9	−43.4	7.9	7.9	13.1
Peer 3	6.4	0.9	5.6	5.5	6.0	6.2	−40.6	9.4	9.4	14.1
Peer 5	5.0	1.7	6.0	5.3	6.5	6.7	−36.5	7.6	7.7	16.0
Peer 2	17.4	0.9	11.9	10.1	11.4	12.2	−53.9	19.3	19.1	16.3

Value at Risk (VaR)
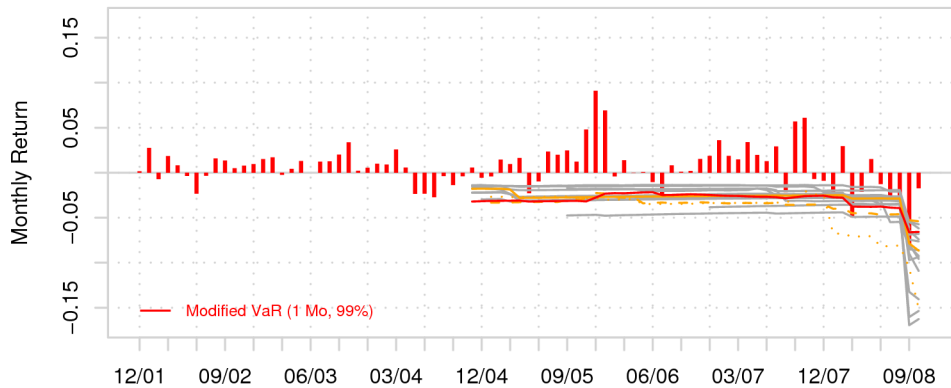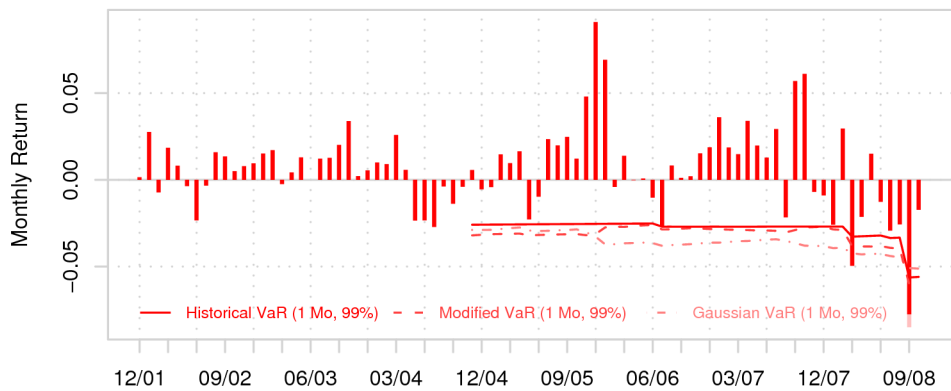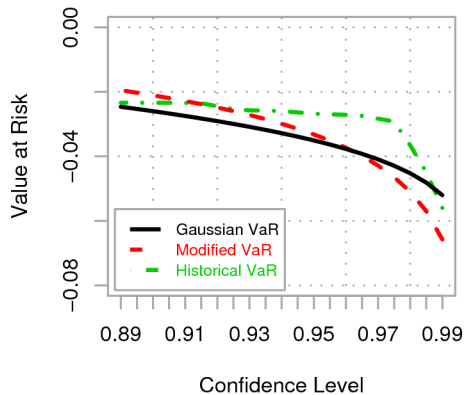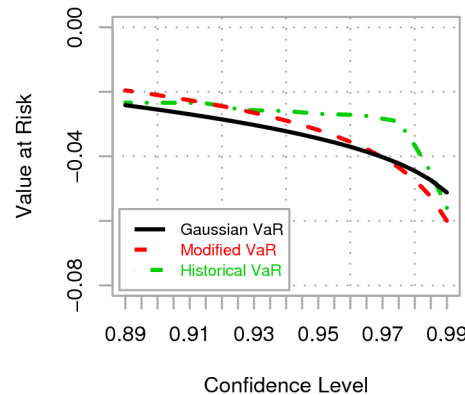
- What level of loss is 'normal'?

- A special case of downside risk, the limitations of mean Value-at-Risk are well covered

- Other single-instrument VaR methods, including Cornish Fisher, Skew-t, Monte Carlo

- Extensions such as Expected Shortfall

- Comparison helps gain some insight into the performance of the different VaR measures. See **chart.VaRSensitivity**

- For example, modified (Cornish Fisher) VaR can be lower than traditional VaR if the asset exhibits positive skewness and low kurtosi

- VaR measures gain extra importance in a portfolio (component) setting
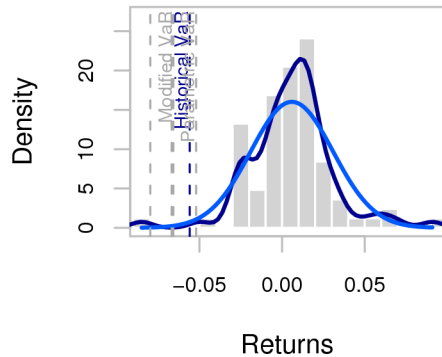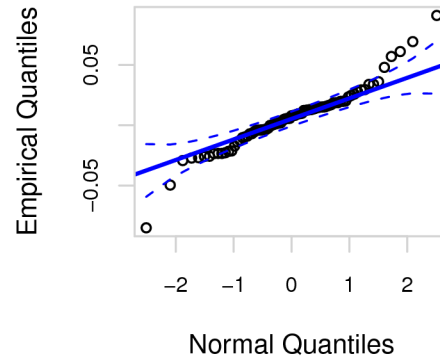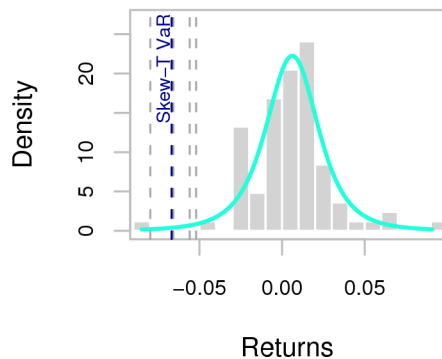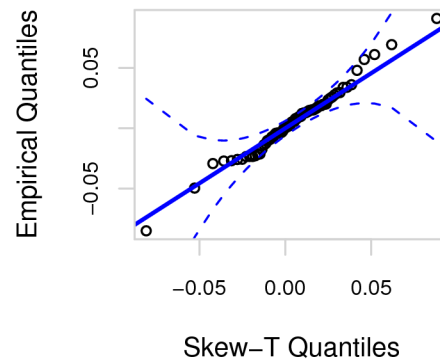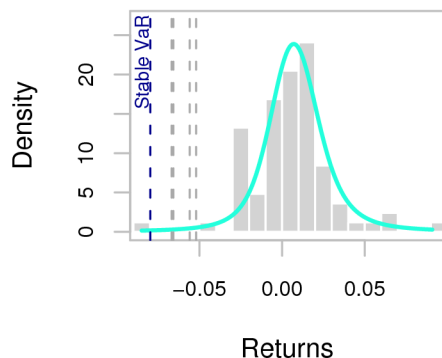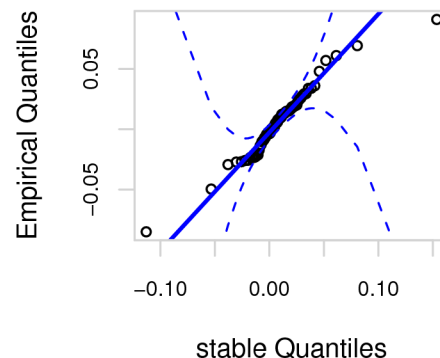
- Can we assume that returns are distributed Normally? Most asset returns exhibit strong deviations from normality

- Histogram visualization depends heavily on bin width, improved by using a kernel density estimator. See **chart.Histogram**

- Quantile-Quantile (QQ) plot compares the data to the theoretical distribution. See **chart.QQPlot**

- Most distributions lack parameters with economic meaning

- Can be useful for expressing what's 'out of bounds' given past performance: When do we know that things may have gone off the rails?

	Beta CoVariance	Beta CoSkewness	Beta CoKurtosis
Fund to EDHEC Conv Arb	0.5981	0.4043	0.3828
Japan CB VR20 TR to EDHEC Conv Arb	0.5596	0.4201	0.4004
Peer 14 to EDHEC Conv Arb	0.5015	0.4632	0.4356
Peer 10 to EDHEC Conv Arb	0.6806	0.6369	0.6205
Peer 11 to EDHEC Conv Arb	0.6525	0.6532	0.6505
Peer 9 to EDHEC Conv Arb	0.7215	0.7101	0.6862
Peer 13 to EDHEC Conv Arb	0.7344	0.7268	0.7376
Peer 12 to EDHEC Conv Arb	0.8072	0.7752	0.7636
Peer 8 to EDHEC Conv Arb	0.8693	0.9543	0.9275
EDHEC Conv Arb to EDHEC Conv Arb	1.0000	1.0000	1.0000
Peer 1 to EDHEC Conv Arb	0.9345	1.0198	1.0376
Peer 4 to EDHEC Conv Arb	1.0077	1.1669	1.1561
Asia.Pac CB VR40 TR to EDHEC Conv Arb	1.0987	1.1708	1.2078
Peer 7 to EDHEC Conv Arb	1.2167	1.2301	1.2276
Peer 6 to EDHEC Conv Arb	1.2029	1.2561	1.2566
Peer 2 to EDHEC Conv Arb	1.7308	1.5463	1.3097
Peer 5 to EDHEC Conv Arb	1.6547	1.7597	1.7518
Peer 3 to EDHEC Conv Arb	1.6708	1.8061	1.7857

- Hasn't been an easily available, standard way for measuring third and fourth order diversification

- Will an asset provide diversification potential in terms of volatility, skewness and kurtosis?

- Martellini (2005): co-moments do not allow the marginal impact of an asset on a portfolio to be directly measured.

- Higher moment betas useful to estimate how much risk will be impacted by adding an asset to a portfolio, in terms of higher moments. See **table.HigherMoments**

- Subject to the quality of assessments of skewness, kurtosis

Fund Rolling 36–Month Style Weights
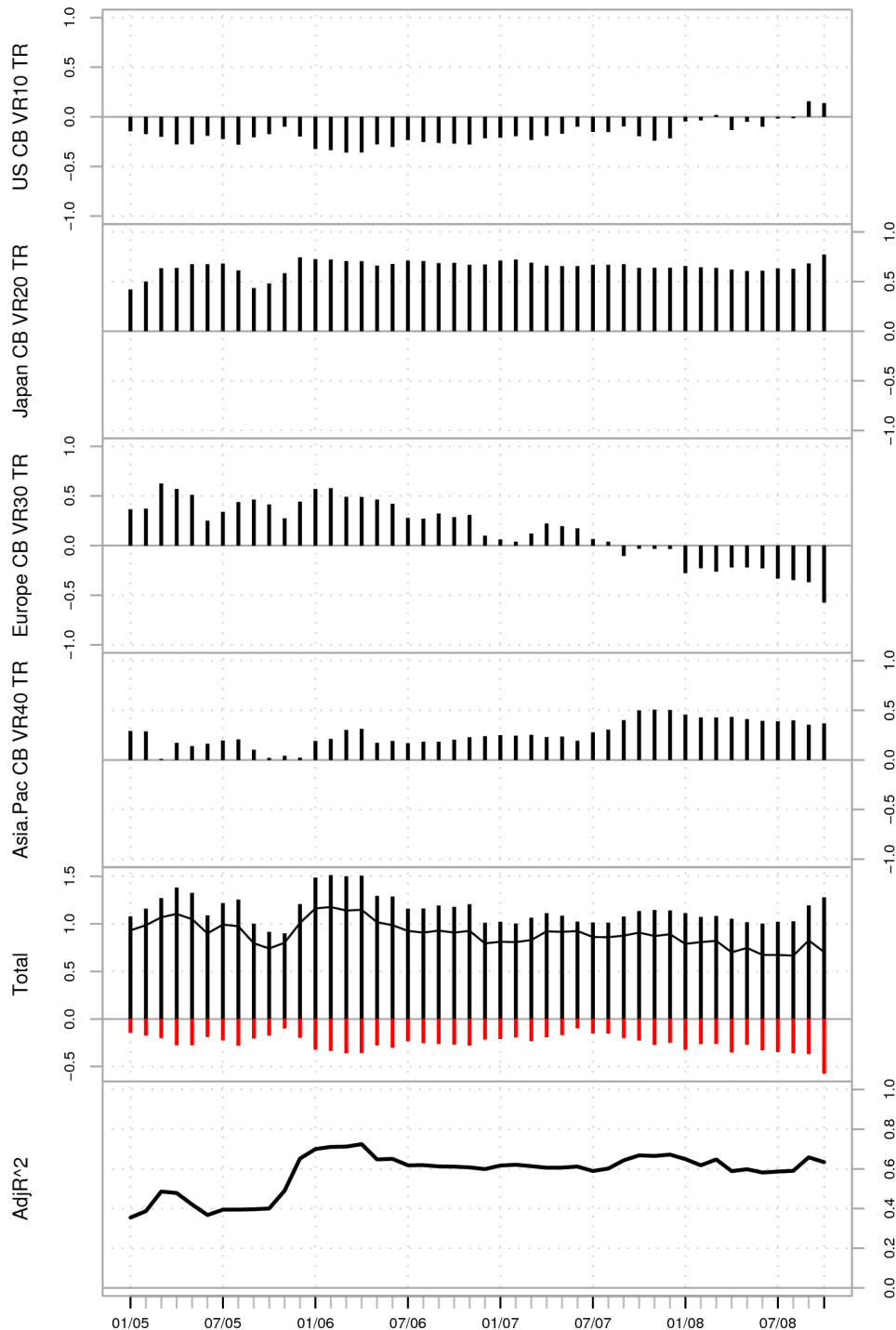
- What risks has the fund taken?

- Can we characterize the source of past returns?

- Sharpe (1992): styles identified may be interpreted as an average of potentially changing exposures over the period

- Rolling-windows may be useful for examining the behavior of a manager's average exposures to asset classes over time

- Very similar to other forms of factor analysis

- See **chart.RollingStyle** and underlying functions

- Difficult to interpret, easy to mis-use

Methods

- The function **chart.Histogram** does this for only a few distributions, so it's good to know how to do it for others.

- Maximum likelihood estimation (MLE) is the most important, widespread, and efficient method used for estimating parameters.

  > x = checkData(managers[, 2, drop = FALSE], na.rm = TRUE, method = "vector")

  > library(sn)

  > fit = st.mle(y = x); s = seq(range(x)[1], range(x)[2], length = 500)

  > fitted.sst = dst(s, location = fit$dp[[1]], scale = fit$dp[[2]], shape = fit$dp[[3]], df=fit$dp[[4]], log = FALSE)

  > chart.Histogram(x, probability=TRUE, main = "Histogram") # hist

  > lines(s, fitted.sst, col = "blue", lwd=2)

  > chart.QQPlot(x, main = "QQ Plot", envelope = 0.95, distribution = "st", location = fit$dp[[1]], scale = fit$dp[[2]], shape = fit$dp[[3]], df = fit$dp[[4]], log = FALSE)

Fit a Distribution Using MLE

- There are two ways to solve for fit parameters. For a known pdf, an explicit formula for the MLE can usually be found.

- The function **fitdistr** in the package MASS uses the closed-form MLEs for the Normal, log-Normal, exponential and Poisson distributions (Venables and Ripley, 2002).

- When a formula is not available, an iterative optimization routine is used to find a solution.

- In R, the **mle** function uses the **optim** optimizer to find the minimum of the negative log-likelihood. The user specifies the negative log-likelihood analytical expression as argument and some starting parameters estimates.

- A quick help.search('mle') will show a number of functions in R that use the MLE approach to fit particular distributions, such as the st.mle function in the package sn.

Modern Portfolio Theory (MPT)

- collection of tools and techniques by which a risk-averse investor may construct an optimal portfolio

- pioneered by Markowitz's 1952 paper *Portfolio Selection*

- also encompasses CAPM, the efficient market hypothesis,

- Can include basically all forms of quantitative portfolio construction and optimization.

Capital Asset Pricing Model (CAPM)

- initially developed by William Sharpe in 1964

- provides a justification for passive or index investing by positing that assets that are not on the efficient frontier will either rise or lower in price until they are on the efficient frontier of the market portfolio

- Extended and evaluated by Sortino, Treynor, Merton, and others

```
>  CAPMtable<-
table.CAPM(Ra=managers[,1:6,drop=FALSE],Rb=managers[,c("S
P500.TR","EDHEC.LS.EQ"),drop=FALSE],scale=12,
rf=managers[,"US.3m.TR",drop=FALSE])
```

- This illustrates why the choice of benchmark is so important: All these managers have very low correlations (Beta) to the S&P, but high correlation to the EDHEC LS Equity index.

- **table.CAPM**() collects most of the interesting CAPM stats

	HAM1 to SP500.TR	HAM1 to EDHEC.LS.EQ	HAM2 to SP500.TR	HAM2 to EDHEC.LS.EQ	HAM3 to SP500.TR	HAM3 to EDHEC.LS.EQ	HAM4 to SP500.TR	HAM4 to EDHEC.LS.EQ	HAM5 to SP500.TR	HAM5 to EDHEC.LS.EQ	HAM6 to SP500.TR	HAM6 to EDHEC.LS.EQ
Alpha	0.0059	0.0031	0.0087	0.0012	0.0059	−0.0016	0.0035	−0.0036	0.0011	−0.0025	0.0088	0.0052
Beta	0.3875	0.7687	0.3489	1.2786	0.5620	1.4618	0.6905	1.6620	0.3099	1.1171	0.3369	1.0662
R−squared	0.4454	0.3659	0.1760	0.5268	0.4353	0.6465	0.3092	0.3733	0.0808	0.1915	0.2839	0.5922
Annualized Alpha	0.0733	0.0384	0.1089	0.0143	0.0734	−0.0196	0.0432	−0.0428	0.0132	−0.0296	0.1112	0.0647
Correlation	0.6674	0.6049	0.4195	0.7258	0.6598	0.8041	0.5560	0.6110	0.2843	0.4376	0.5328	0.7696
Correlation p−value	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0122	1e−04	0.0000	0.0000
Tracking Error	0.1032	0.0644	0.1957	0.1122	0.1412	0.0482	0.0739	0.0203	0.0407	0.0654	0.1791	0.1547
Active Premium	0.0412	0.0202	0.0698	0.0330	0.0492	0.0104	0.0170	−0.0103	0.0141	−0.0363	0.0880	0.0701
Information Ratio	0.3991	0.3136	0.3567	0.2942	0.3482	0.2161	0.2306	−0.5064	0.3460	−0.5557	0.4913	0.4533
Treynor Ratio	0.2491	0.1247	0.3616	0.0850	0.1859	0.0589	0.1048	0.0394	−5e−04	−1e−04	0.3674	0.1161

- **CAPM.alpha**: the degree to which the asset's returns are not explained by the return that could be captured from the market. Conversely,

- **CAPM.beta**: describes the portions of the returns of the asset that could be directly attributed to the returns of a passive investment in the benchmark asset (via linear model regression)

- **TreynorRatio**: ratio of asset's Excess Return to Beta β of the benchmark

- **ActivePremium**: investment's annualized return minus the benchmark's annualized return

- **TrackingError**: a measure of the unexplained portion of performance relative to a benchmark

- **InformationRatio**: ActivePremium/TrackingError. Information Ratio may be used to rank investments in a relative fashion
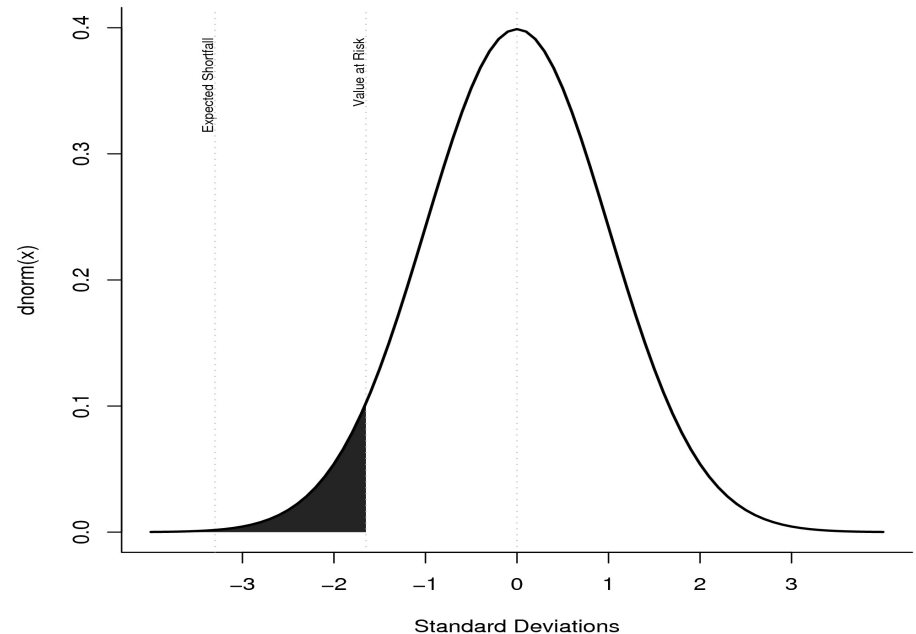
Downside Risk

- Most investors are properly concerned with both reward and risk. See **table.DownsideRisk**

- Standard Deviation or Volatility is often used as the primary/ first measure of risk, but it is symmetrical

- Markowitz proposed using **SemiDeviation** to measure only downside volatility

- As already covered, most investments incorporate significant variations from normality in either **skewness** or **kurtosis**

- Sortino and Price (2004) propose incorporating a minimum acceptable return (MAR) into the measure which results in **DownsideDeviation** and is used in **UpsidePotentialRatio**

- The other very widely used method of evaluating downside risk is through analysis of drawdowns, provided in **findDrawdowns**, **sortDrawdowns**, **chart.Drawdown**, **maxDrawdown**, and **UpDownRatios**

- Sharpe first suggested a ratio of return to risk (measured by annualized [excess] return over annualized standard deviation) which is implemented as **SharpeRatio**

- William Sharpe now recommends **InformationRatio** preferentially to the original Sharpe Ratio

- Many other authors have modified the risk measure in use to create similar ratios

- **SortinoRatio** return over downside deviation

- **UpsidePotentialRatio** improvement of the Sortino ratio that measures upside return in good runs over losses in drawdown periods

- Several other modified Sharpe ratios have been proposed, and are implemented in PerformanceAnalytics, including using any of Gaussian and Cornish Fisher modified VaR and ES measures as the risk measure in parameters to **SharpeRatio.modified**

- Value at Risk (VaR) has become a required standard risk measure recognized by Basel II and MiFID

- PerformanceAnalytics 1.0 has a **VaR** function with standardized parameters for historical, Gaussian, kernel, and Cornish Fisher VaR.

- Wrapper will probably be extended to include MonteCarlo, Skew Student-t, and generalized Pareto VaR estimators soon
- Key parameters are R{the series}, p{probability}, and the method desired.
- Also allows moments to be passed in to allow for different estimators or optimization
- Also allows for Marginal and Component VaR calculations

- *Marginal VaR* is the change in the portfolio VaR attributable to a change in position at the margin.

  - Implemented in the **VaR** wrapper or as **VaR.Marginal**

  - Not particularly useful, except in special cases for rebalancing similar instruments

- *Component VaR* is the risk contribution of each instrument to the risk of the whole portfolio

Expected Shortfall (ES)

- *Conditional Value at Risk (CvaR)* also known as *Expected Shortfall (ES)* is the mean expected loss when the loss exceeds the VaR

- We have wrapped **ES** in a manner similar to **VaR** and provide both single-instrument and component versions for Gaussian, Historical, Kernel, and Cornish Fisher Expected Shortfall

- Beyond VaR (**VaR.Beyond**) is a related measure of mean expected tail loss that adds VaR and ES

	HAM1	HAM2	HAM3	HAM4	HAM5	HAM6
Gaussian ES	–0.0106	–0.0128	–0.0112	–0.0091	–0.0023	–0.0115
Gaussian Beyond VaR	–0.0800	–0.1120	–0.1093	–0.1438	–0.1090	–0.0784
Modified ES	–0.0106	–0.0115	–0.0109	–0.0093	–0.0017	–0.0115
Modified Beyond VaR	–0.0772	–0.1420	–0.1431	–0.1353	–0.1316	–0.0713

- Some investors, risk managers, or regulators will ask for confidence of 99.99% or even higher.

- This means:
  - Once in 85 years for monthly data
  - Once in 20 years for weekly data
  - Once in 4 years for daily data

- These kinds of risk confidence probabilities only make sense if you have a LOT of data, or for high-frequency data over short horizons (minutes to maybe an hour)

- High confidence levels also assume a stable series, probably a bad assumption over the timeframes above.

- Developing these confidence levels by extending/expanding your time series through simulation may work

- Beware of hidden distributional assumptions

VaR Sensitivity of HAM1

VaR Sensitivity of HAM2
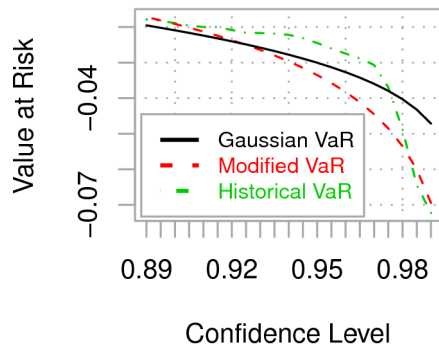
VaR Sensitivity of HAM3

VaR Sensitivity of HAM4

- In addition to "false confidence", you also need to be aware of when the risk measure you're using doesn't fit the data well.
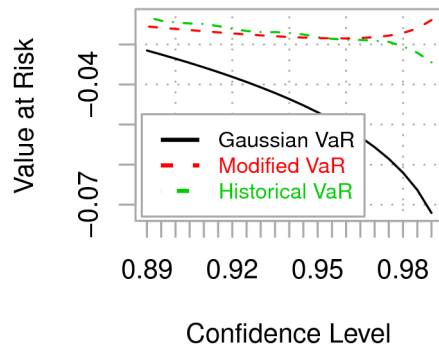
>layout(matrix(c(1,2,3,4),nrow=2,ncol=2,byrow=TRUE))

>apply(managers[,1:4,drop=FALSE],2,chart.VaRSensitivity)

- Note that Cornish Fisher VaR generally more closely mirrors historical realizations

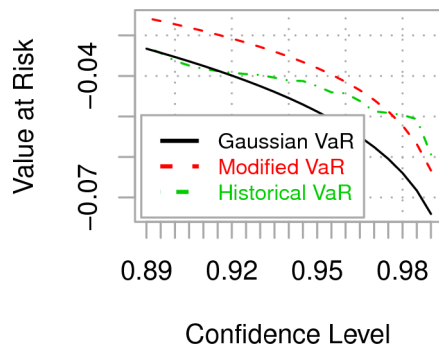- Note where HAM2 breaks down completely at high probabilities

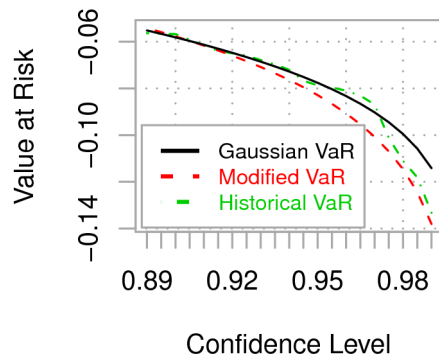- Out-of-sample predictions of return or risk may be adversely affected by autocorrelation or outliers in the data

- Geltner and Getmansky have proposed different methods of dealing with autocorrelation implemented as **Return.Geltner** and **SmoothingIndex**

- We propose and test a method in our 2008 JoR paper for a robust method of cleaning returns to improve out of sample risk predictions.

  - identify the returns that are outside the confidence threshold desired for the risk measure (e.g. 95% or 99%)

  - reduce the magnitude of those outlying returns if they are outside the range of the other returns as identified by the Mahalanobis distance

  - reduction in magnitude keeps ranked magnitude intact (the largest loss/gain will remain the largest loss/gain, even after smoothing)

  - see Boudt,Peterson,Croux (2008) for details or Return.clean
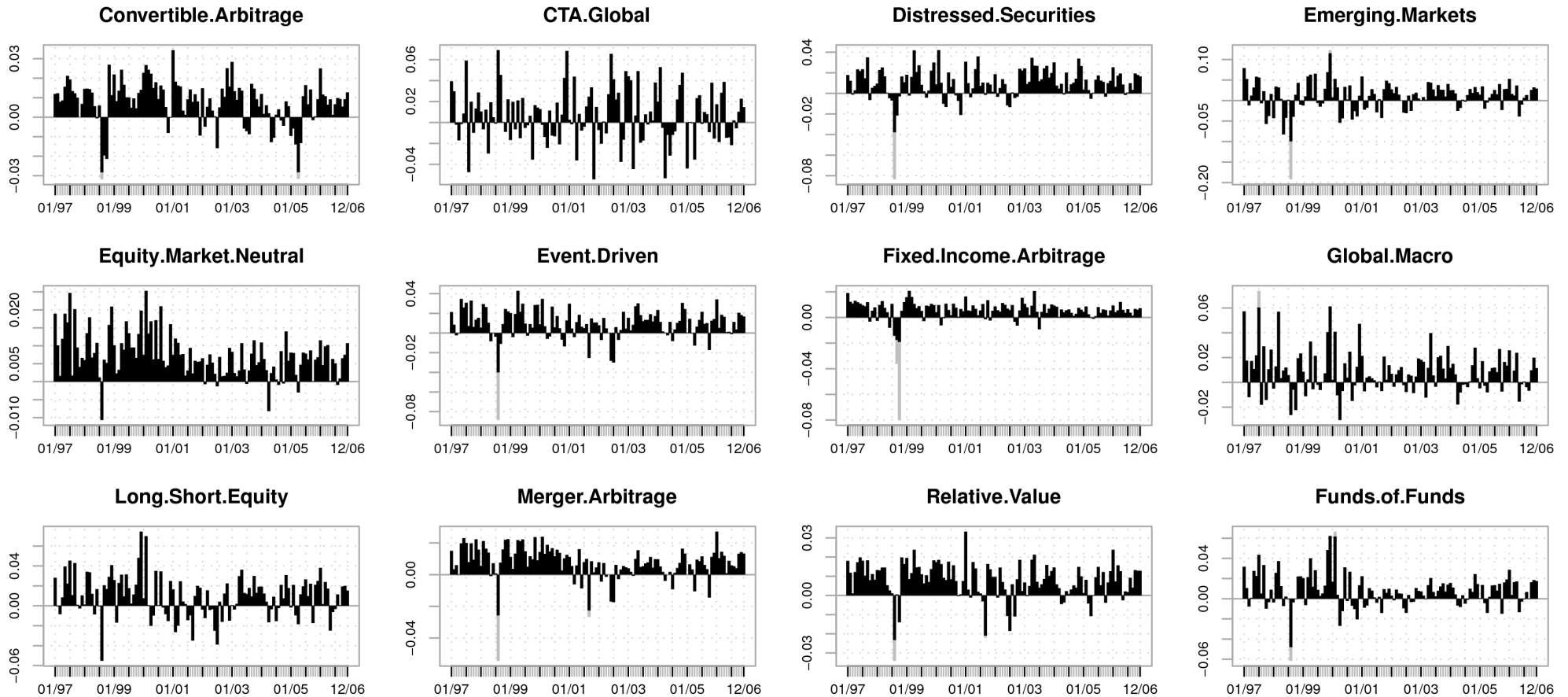
- Note the smoothing of outlying returns around the Russian financial crisis.

- Robust smoothing improves out of sample risk predictions

- 'edhec' data series cleaned by **Return.clean** via the function **chart.barVaR**

- Style analysis is a returns-based analysis used to determine a fund's exposures to the changes in returns of major asset classes.

- **chart.RollingStyle** calculates and displays effective weights in rolling windows through time. **style.fit** manages the calculation of the weights by method. **style.QPfit** calculates the specific constraint case that requires quadratic programming.

- The "unconstrained" method implements a multiple regression. "Normalized" is the same, but the coefficients sum to 1. The "constrained" adds that the coefficients must be between 0 and 1. That requires a quadratic programming algorithm, using **solve.QP** from the 'quadprog' package.

- Uses R^2 for variable selection. Alternatively, we have an implementation that uses AIC as a criterion-based method for selecting style weights. Email us if you're interested.

```
>data(edhec)

>chart.RollingStyle(managers[,2,drop=FALSE],edhec[,1:11],
method="constrained", leverage=FALSE, width=36, cex.legend
= .7, colorset=rainbow12equal, las=1)
```

- Apologies in advance for the stacked bar chart.

- A much better alternative is to graph each factor weight separately so that the values can be read easily and are easier to compare

Further Research

- Performance attribution and Portfolio Analytics

- Risk Budgeting

  – Applications of Component Risk to large portfolios

- Historical portfolio framework in R

  – Transactions and positions with profit and loss

  – Instrument properties and model

  – Interfaces to data provider packages

- Practical Bayesian applications

- Anything YOU want to work with us on

Thank You for your attention

How

- There are several ways to install packages, but the preferable way to install them is through the CRAN repository.

- Within an R session, you can type:

    > install.packages('PerformanceAnalytics', dep=TRUE)

- Once the package is installed, you need to load it into the current workspace to use it. Use

    > library(PerformanceAnalytics)

  to load the package.

- Alternatively, under Windows, use the 'Package' menu and select 'Library.'

- **How do I read data from an Excel spreadsheet?**

  - Export the data you want from Excel into a comma-delimited file (usually saved with a .csv suffix) and use read.csv

  - Separate data using commas and with a header above each column.

  - For time series data, the date is usually found in the first column in YYYY-MM-DD format

  - Look at documentation for **Return.read** and **read.csv**

- **How do I read data from Internet sources?**

  - There are a variety of ways, but we recommend the quantmod package, see **?getSymbols**

  - Handles a variety of Internet sources including Yahoo! Finance, FRED, Google Finance, and Oanda.

  - Provides price aggregation from higher to lower timescales, see **? to.period**

  - Also spectacular price charts

- Example data for this presentation is included in the PerformanceAnalytics package. To load the data, type:

  > data(managers)

- 'managers' is a data frame that contains columns of (disguised) monthly returns for six equity long-short hedge funds, plus

  - the EDHEC Long-Short Equity hedge fund index,

  - the S&P 500 total returns,

  - a total return series for the US Treasury 10-year bond and

  - a total return series for the US Treasury 3-month bill.

- Monthly returns for all series end in December 2006 and begin at different periods starting from January 1996.

- Always choose the least complex date/time class that will accomplish the task.

  - POSIX classes support times and dates including time zones and daylight savings time. They are mostly useful for intraday and tick data, such as currency data where data alignment is an issue.

  - Date is a class in base that supports dates without times.

- The Date class provides a seq.Date function:

  ```
  > seq(as.Date("1999-12-01"),length.out=12, by="1 month")
  ```

  ```
  > seq(ISOdate(2000,1,31), by = "month", length = 4)
  ```

  ```
  > seq(as.Date("2000-01-01"),length.out=120, by="1 month")-1
  ```

- The zoo package provides general support for ordered observations, independent of a particular index class.

  - Any index will work, including the indexes provided by ts, its, irts and timeSeries.

  - Provides careful backwards compatibility with robust as.* methods and proper use of reclass.

- The xts package extends zoo by enforcing a time-based index but allowing for multiple timescales.

- xts is very fast

- As an extension to zoo, xts offers relatively lossless coversion to and from the myriad of time-series classes in R. Calling reclass turns it back into the object's original form with everything intact.

- Provides a unique method of date subsetting.

    > sample.xts['2007'] # all of 2007

    > sample.xts['2007-03::'] # March 2007 to the end of the data set

    > sample.xts['2007-03::2007'] # March 2007 to the end of 2007

    > sample.xts['::2007'] # the beginning of the data through 2007

    > sample.xts['2007-01-03'] # just the 3rd of January 2007

    – This format extends all the way down to time - e.g. '1999-01-05 08:00::1999-01-05 08:35:23'.

- Most market data is prices, but almost the entire body of methods for performance and risk analysis requires returns.

- Prices can be easily converted to simple or compound (logarithmic) returns using **Return.calculate**

- The single-period arithmetic return, or simple return, can be calculated (using quantmod to get the data):

  ```
  > getSymbols("SPY")

  > Return.calculate(SPY$SPY.Close,method="simple")
  ```

- Simple returns, cannot be added together. Simple returns are most often the correct input for performance or risk analysis.

- The natural logarithm of the simple return of an asset is referred to as the continuously compounded return, or log return, and may be added together to get a total return.

  ```
  > Return.calculate(SPY$SPY.Close,method="compound")
  ```

- This is an example of how the data was set up for all of the examples shown in the first half. All of the following code snippets reference these variables:

  > manager.column = 1; peer.columns = 2:6; index.columns = 7:8; manager.color = "red"; peer.color = "darkgray"; index.color = "orange"

  > colorset = c(rep(manager.color,length(manager.column)), rep(index.color, length(index.columns)), rep(peer.color,length(peer.columns)))

  > linetypes = c(rep(1, length(manager.column)), 1:length(index.columns), rep(1, length(peer.columns)))

  > dottypes = c(rep(16, length(manager.column)), closedsymbols[1:length(index.columns)], rep(1, length(peer.columns)))

- Charts in PerformanceAnalytics are designed to be compatible with traditional R graphics such as plot.

- For example, line widths can be changed by setting lwd = 2, or line types can be changed using lty = 3, etc.

- See ?par and ?plot for the full list of attributes that can be changed.

- For example, a typical time series chart for cumulative returns might look like:

  ```
  > chart.CumReturns(managers[,c(manager.column,  index.columns,
    peer.columns), drop = FALSE], main = "Cumulative Returns",
    legend.loc = "topleft", event.lines = risk.dates, event.labels =
    risk.labels, ylog = TRUE, wealth.index = TRUE, colorset = colorset,
    lwd = 2)
  ```

- PA includes sets of color palattes designed to create readable (but appealing) line and bar graphs, in three categories:

  - Focus: to provide visual focus to the data graphed first. Includes redfocus, bluefocus, and greenfocus. One color and grays.

  - Equal-weighted: For when all of the data should be observed and distinguishable on a line graph. The different numbers in the name indicate the number of colors in the set. See rainbow12equal, rich12equal, tim12equal, dark6equal, set8equal.

  - Monochrome: same color, different value. Includes greenmono , bluemono, redmono, gray8mono and gray6mono.

  - These are just lists of strings that contain the RGB codes of each color. Create your own colorsets with RColorBrewer:::brewer.pal, rainbow, gplots:::rich.colors, fields:::tim.colors, etc.

- Similarly, there are a few sets of grouped symbols for scatter charts:

  - opensymbols, closedsymbols, fillsymbols, linesymbols, and allsymbols.

- Default legends are easy to use by setting:
  - legend.loc = any of: "bottomright", "bottom", "bottomleft", "left", "topleft", "top", "topright", "right" and "center"

- Or custom legends can be set up after a plot is made. In this case, we're shortening the list of labels:

  > legend.colorset = c(rep(manager.color, length(manager.column)), peer.color, rep(index.color, length(index.columns)))

  > legend.linetypes = c(rep(1, length(manager.column)), 1, 1:length(index.columns))

  > legend.dottypes = c(rep(16, length(manager.column)),1, closedsymbols[1:length(index.columns)])

  > legendnames = c(colnames(x[, manager.column, drop = FALSE]),"Peer group", colnames(x[, index.columns, drop = FALSE]))

  > legend("topleft", inset = 0.02, text.col = legend.colorset, col = legend.colorset, cex = .8, border.col = "gray", lwd = 2, bg = "white", legend = legendnames, lty=legend.linetypes, pch = legend.dottypes, pt.bg="white", pt.lwd = "1", merge = FALSE, pt.cex = 1.25)

- Create a data.frame arranged how you want, either by hand or using one of the table.* functions:

  ```
  > w = t(table.Autocorrelation(x[, c(manager.column, peer.columns,
    index.columns), drop = FALSE]))
  ```

- Use format.df to format contents of a data.frame

  - Found in the 'Hmisc' package

  ```
  > library(Hmisc)
  ```

  ```
  > x = format.df(w, na.blank=T, rdec=c(rep(4,dim(w)[1])),
    col.just=rep("nc",dim(w)[2]))
  ```

- Use textplot to display text output in any graphics device.

  - Originally from 'gplots' package, but we've added text wrapping for long labels and a few other features.

  ```
  > textplot(x,  cex = 1, rmar = 0.65, cmar = 1.0, mar=c(0,0,0,0), halign =
    "center", valign = "center", wrap.rownames=50)
  ```

- How to use layout for charts:

```
> layout(matrix(c(1,1,2,2,3,3,4,5),
    nrow=4, ncol=2, byrow=TRUE),
    height=c(3,1.25,1.75,3), width=1)
```

```
> layout.show(5)
```

- Unfortunately, you can't nest layouts. Some charts that use layout can not then, in turn, be used with layout.

- Margins for plots can be changed using par in each frame

  - mar: a numerical vector of the form c(bottom, left, top, right) which gives the number of lines of margin to be specified on the four sides of the plot. The default is c(5, 4, 4, 2) + 0.1

```
> par(mar=c(1,4,4,2))
```

- **Writing to PDF files**

  > pdf(file=”filename.pdf”, height=10, width=7.5, paper="letter")

  …

  > dev.off()

- **… or Encapsulated Postscript**

  – Works with Word, Latex

  > postscript(file="filename.eps", width=5.25, height=8, paper="special", horizontal=FALSE, onefile=FALSE)

  …

  > dev.off()

- **Copying from a device**

  – May not copy accurately

  > dev.copy2eps()