# Can you do better than cap-weighted equity benchmarks?

Guy Yollin
Principal Consultant, r-programming.org
Visiting Lecturer, University of Washington

Krishna Kumar
Financial Consultant

# Legal Disclaimer

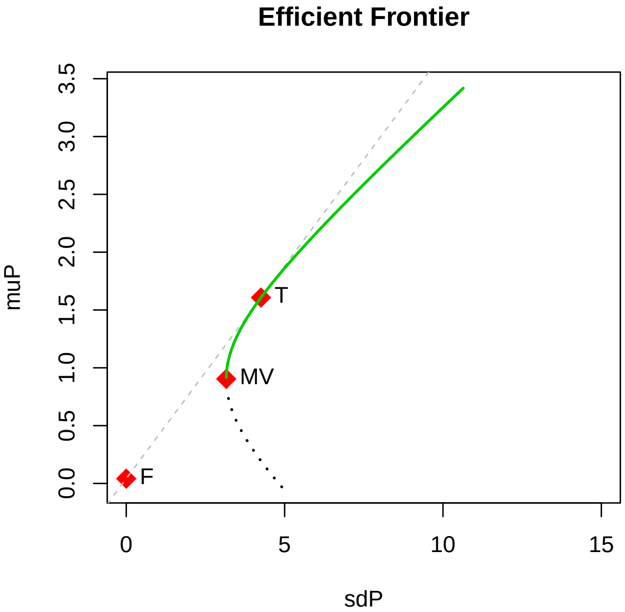
- This presentation is for informational purposes
- This presentation should not be construed as a solicitation or offering of investment services
- The presentation does not intend to provide investment advice
- The information in this presentation should not be construed as an offer to buy or sell, or the solicitation of an offer to buy or sell any security, or as a recommendation or advice about the purchase or sale of any security
- The presenter(s) shall not shall be liable for any errors or inaccuracies in the information presented
- There are no warranties, expressed or implied, as to accuracy, completeness, or results obtained from any information presented

INVESTING ALWAYS INVOLVES RISK

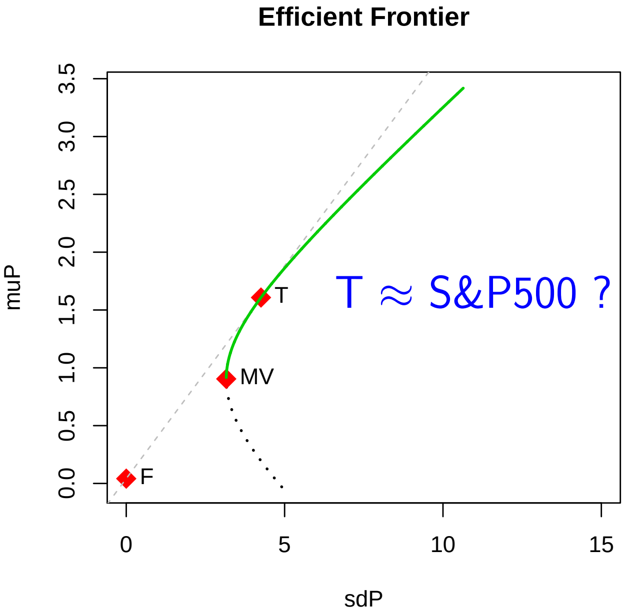
# Outline

# Outline

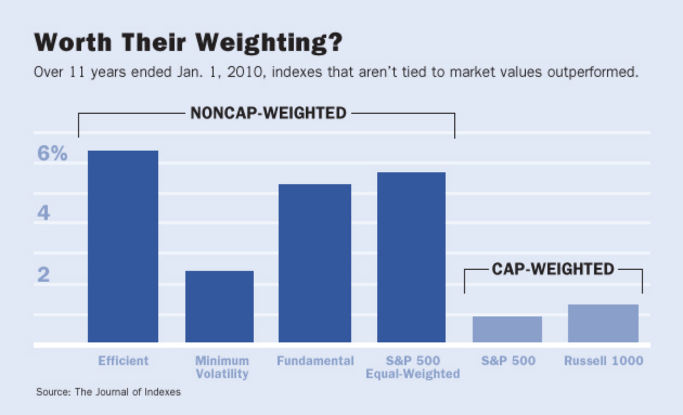
# The tangency portfolio



# The tangency portfolio



# Is Your Index Fund Broken?




Jack Hough, SmartMoney, "Is Your Index Fund Broken?", January 31, 2011

# A New Idea?

**The efficient market inefficiency of capitalization-weighted stock portfolios**

*"Matching the market is an inefficient investment strategy."*

*Robert A. Haugen and Nardin L. Baker*

---

Haugen and Baker, Journal of Portfolio Management, "The efficient market inefficiency of capitalization-weighted stock portfolios", Spring 1991

# Motivation for research

Efficient Indexation

- maximize Sharpe ratio

$$w^* = \arg\max_{w} \frac{w'\mu}{\sqrt{w'\Sigma w}}$$

- covariance matrix
  - derived from principal component analysis (PCA)
- expected returns
  - form deciles by downside risk
  - expected return equals mean of each decile

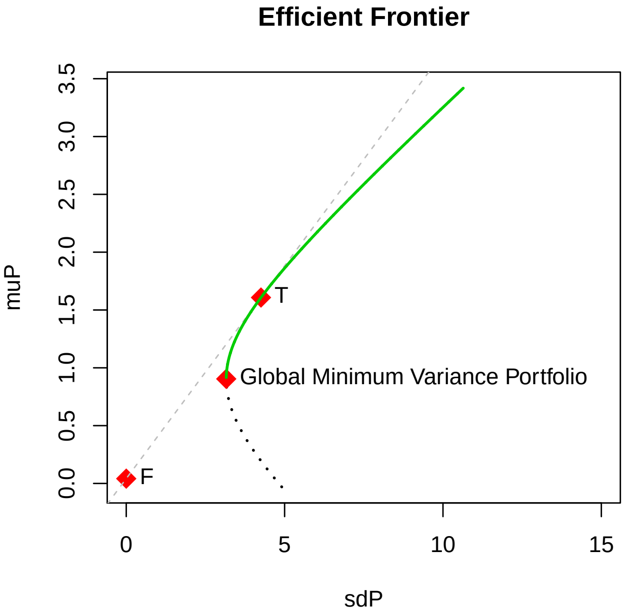

Amenc, Goltz, Martellini, "Efficient Indexation: An Alternative to Cap-Weighted Indices", January 2010

# Research project

- Goal
  - Compare performance of alternative index constructions using S&P 500 constituents
- Methodology
  - use a rolling 2-year window of current constituent returns and re-balance at the start of each month
  - generate 48-months of out-of-sample index returns (Jan-2007 to Dec-2010)
  - S&P 500 returns were calculated using constituent weights (apples-to-apples comparisons without factoring in transaction costs)
- Constraint
  - positive weights (max of 25%)
- Focus of research
  - minimum risk (minimum variance and minimum CVaR) portfolios

# Outline

# Global minimum variance portfolio



# M-V optimization and Quadratic Programming

general QP problem

$$min_b \quad \frac{1}{2}\mathbf{b}^\mathbf{T}\mathbf{D}\mathbf{b} - \mathbf{b}^\mathbf{T}\mathbf{d}$$

$$s.t. \quad \mathbf{A}^\mathbf{T}\mathbf{b} \geq \mathbf{b_0}$$

$$\mathbf{b} \geq 0$$

mean-variance portfolio optimization

$$min_b \quad \boldsymbol{\omega}^T \boldsymbol{\Sigma} \boldsymbol{\omega}$$

$$s.t. \quad \boldsymbol{\omega}^T \boldsymbol{\mu} = \mu_p$$

$$\boldsymbol{\omega}^T \mathbf{1} = 1$$

$$\omega_{min} \geq \omega_i \geq \omega_{max}$$

**R Code: the solve.QP function**

```
> library(quadprog)
> args(solve.QP)

function (Dmat, dvec, Amat, bvec, meq = 0, factorized = FALSE)
NULL
```

objective function assignments: $2\boldsymbol{\Sigma} \rightarrow \mathbf{D} \quad \boldsymbol{\omega} \rightarrow \mathbf{b} \quad \mathbf{0} \rightarrow \mathbf{d}$

# Factor models for asset returns

The general form of a factor model for asset returns is:

$$R_{j,t} = \beta_{0,j} + \beta_{1,j} F_{1,t} + \cdots + \beta_{p,j} F_{p,t} + \epsilon_{j,t}$$

where

$R_{j,t}$ is either return or excess return on the jth asset at time t

$F_{1,t}, \ldots, F_{p,t}$ are factors (aka risk factors) at time t

$\epsilon_{1,t}, \ldots, \epsilon_{n,t}$ are uncorrelated, mean-zero, unique risks

The factor model in matrix form is:

$$\mathbf{R_t} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}^T \mathbf{F_t} + \boldsymbol{\epsilon}_t$$

# Returns covariance matrix

Given the following covariance matrices:

$$\boldsymbol{\Sigma}_{\epsilon} = \begin{pmatrix} \sigma^2_{\epsilon,1} & \cdots & 0 \\ \vdots & \sigma^2_{\epsilon,j} & \vdots \\ 0 & \cdots & \sigma^2_{\epsilon,n} \end{pmatrix}$$

$$\boldsymbol{\Sigma}_F = p \times p \text{ covariance matrix of } (F_t)$$

The returns covariance matrix is:

$$\boldsymbol{\Sigma}_R = \boldsymbol{\beta}^T \boldsymbol{\Sigma}_F \boldsymbol{\beta} + \boldsymbol{\Sigma}_{\epsilon}$$

# Covariance matrix estimation

- Estimating the covariance matrix based on a factor model is a bias-versus-variance trade-off
  - sample covariance matrix is unbiased but may have *significant* estimation error
  - estimating the covariance matrix via a factor model may be biased but also may significantly reduce estimation error by significantly reducing the number of estimates

- Sample covariance matrix for n-assets
  - $n(n+1)/2$ estimations
  - for 500 assets, 125,250 estimates are required

- Covariance matrix with n-assets and a factor model with p-factors
  - $np + n + p^2$ estimations
  - for 500 assets and 10 factors, 5,600 estimates are required

# Industry factor model

Model background

- Sheikh, "Barra's Risk Models", 1995

Response

- daily equity returns

Explanatory variables

- company industry classification

Model details

- Example 103, Zivot, "Modeling Financial Time Series with S-PLUS, 2nd Edition", 2005
  `http://faculty.washington.edu/ezivot/book/Ch15.factorExamples2ndEdition.ssc`

# Cross-sectional factor models

Differences between time-series factor models and cross-sectional factor models:

| **Model type** | **Assets** | **Time Periods** | **Factors** | **Betas** |
|---|---|---|---|---|
| time-series | one asset at a time | all time periods | known | estimated |
| cross-section | all assets | one period at a time | estimated | known |

Cross-sectional factor model for the jth asset at some fixed t:

$$R_j = \beta_0 + \beta_1 F_{1,j} + \cdots + \beta_p F_{p,j} + \epsilon_j$$

# Industry factor model

General industry factor model has the following form:

$$R_j = \beta_1 F_{1,j} + \beta_2 F_{2,j} + \cdots + \beta_p F_{p,j} + \epsilon_j$$

$$\beta_i = \begin{cases} 1, & \text{if asset j in industry i} \\ 0, & \text{if asset j not in industry i} \end{cases}$$

- Factor realizations represent a weighted average return in time period t of all of the asset returns for companies operating in industry j
- S&P Sector GICS codes for 10 sectors (10 sectors):

| | | | | |
|---|---|---|---|---|
| energy | materials | industrial | discretionary | staples |
| health | financial | info tech | telecom | utilities |

# Statistical factor models

Recall the general form of a factor model:

$$\mathbf{R_t} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}^T \mathbf{F_t} + \boldsymbol{\epsilon}_t$$

- In statistical factor models:
  - factor realizations are not directly observable
  - no external knowledge of betas (as in cross-sectional models)
  - factor realizations and betas must be extracted from the returns data using statistical methods
- Principal component analysis - eigen decomposition of covariance matrix

# PCA statistical factor model

Model background

- "Modeling Financial Time Series with S-PLUS, 2nd Edition", 2005

Response

- daily equity returns

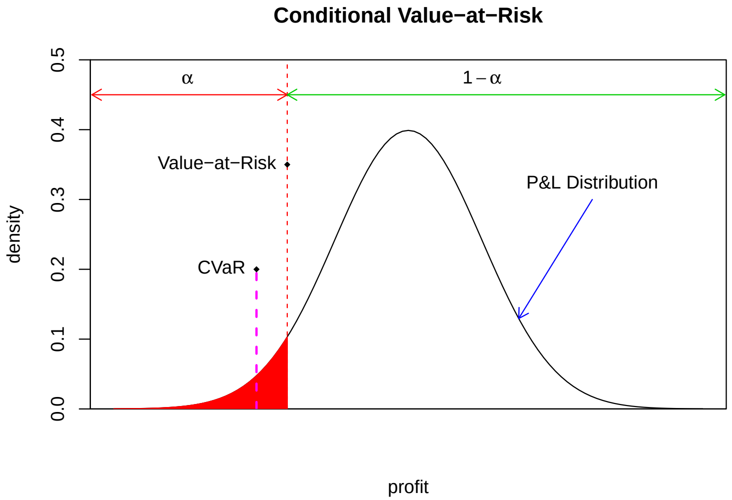
Explanatory variables

- principal components

Model details

- Example 112, Zivot, "Modeling Financial Time Series with S-PLUS, 2nd Edition", 2005
  `http://faculty.washington.edu/ezivot/book/Ch15.factorExamples2ndEdition.ssc`

# Conditional Value-at-Risk



# CVaR Optimization via Linear Programming

It can be shown that minimizing the CVaR of a portfolio is a linear programming problem that can be carried out with a general-purpose LP solver†

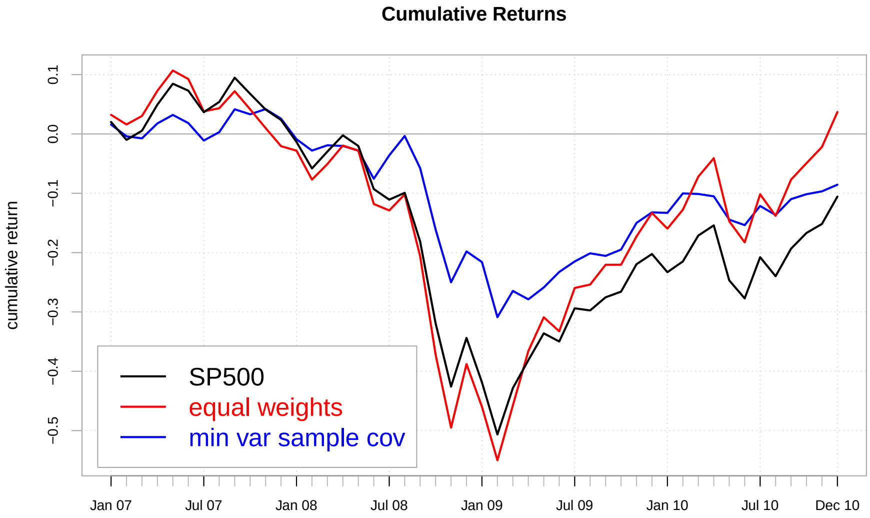
**R Code: the Rglpk_solve_LP**

```
> library(Rglpk)

Using the GLPK callable library version 4.42

> args(Rglpk_solve_LP)

function (obj, mat, dir, rhs, types = NULL, max = FALSE, bounds = NULL,
    verbose = FALSE)
NULL
```

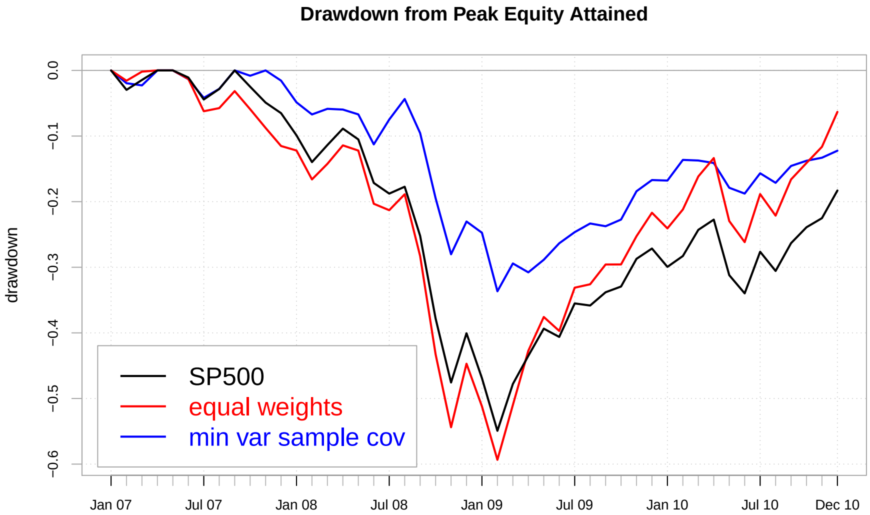
†Yollin, "R Tools for Portfolio Optimization", R/Finance 2009

# Outline

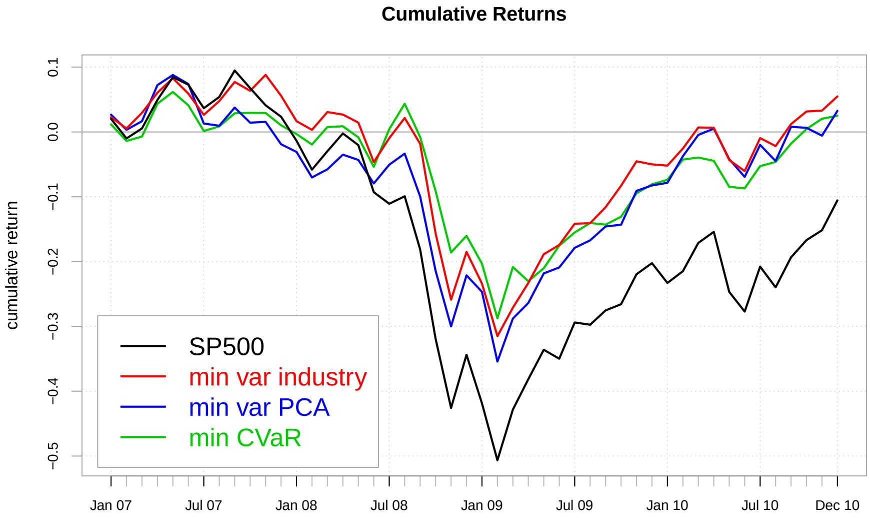
# Cumulative return comparisons



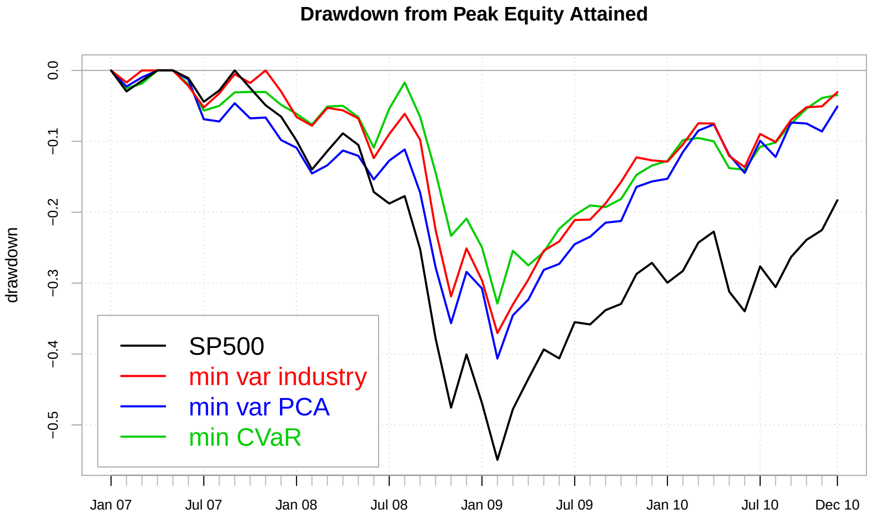
# Cumulative return comparisons



# Cumulative return comparisons



# Cumulative return comparisons



# Summary

| | SP500 | minVaRSample | minVarIndustry | minVarPCA | minCVaR |
|---|---|---|---|---|---|
| Cumulative Return | -0.106 | -0.086 | 0.055 | 0.032 | 0.025 |
| Annualized Return | -0.028 | -0.022 | 0.013 | 0.008 | 0.006 |
| Annualized StdDev | 0.241 | 0.138 | 0.161 | 0.174 | 0.139 |
| Conditional VaR | -0.159 | -0.105 | -0.118 | -0.126 | -0.100 |
| Max DrawDown | 0.549 | 0.337 | 0.370 | 0.406 | 0.329 |

- all minimum variance portfolios and the minimum CVaR portfolio outperformed the S&P 500 Index during the testing period
  - higher annualized return
  - lower annualized volatility
  - smaller conditional value-at-risk
  - smaller maximum drawdown
- returns are difficult (impossible) to forecast and these techniques don't require them

Can you do better than cap-weighted equity benchmarks? Maybe!

# Outline

# Special thanks

SunGard Financial Systems

- Historical S&P 500 constituent weights
- Historical stock prices

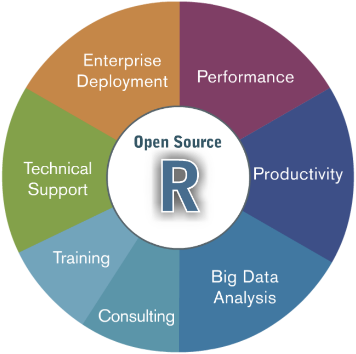
# Special thanks

Revolution Analytics

- Revolution R Enterprise and RevoScaleR

**R is Ready for Business™**

**with Revolution R Enterprise**



# Q & A

- Questions and comments
- Contacting the Presenters
  - Guy Yollin
    - `http://www.r-programming.org`
    - gyollin@r-programming.org
  - Krishna Kumar
    - kk2250@gmail.com