QF623: Portfolio management I
Portfolio construction I

Main drawback of MV portfolios
Quote: Michaud (1989)

"The indifference to many investment professionals to mean-variance optimization technology, despite its theoretical appeal, is understandable in many cases. The major problem with mean-variance optimization is its tendency to **maximize the effects of errors in input assumptions**. **Unconstrained** MV optimization can yield results that are **inferior to those of simple equal-weighting schemes**"

Limitations of MV portfolios
Sensitivity of input parameters - Example

$$\mu_1 = \mu_2 = 8\%, \mu_3 = 5\%$$
$$\sigma_1 = 20\%, \sigma_2 = 21\%, \sigma_3 = 10\%, \rho = 80\%$$
$$x^* = (38.3\%, 20.2\%, 41.5\%)$$

ρ		70%	90%		90%	
σ_2				18%	18%	
μ_1						9%
x_1	38.3	38.3	44.6	13.7	−8.0	60.6
x_2	20.2	25.9	8.9	56.1	74.1	−5.4
x_3	41.5	35.8	46.5	30.2	34.0	44.8

Error maximization

- Portfolio optimizers amplify the effect of estimation error.
  - Overweight securities that have unusually high estimated risk premia and/ or low estimated risk
  - Underweight securities that have unusually low estimated risk premia and/or high estimated risk.
  - Extreme estimates are more likely to contain estimation error.

Limitations of MV portfolios
Jobson and Korkie (1980) experiment

- Objective
  - To ask how reliable plug-in estimates of mean-variance efficient portfolio weights are for given sample sizes
- Experiment
  - Compute the true mean-variance efficient frontier from historical means and covariance matrix
  - Simulate independent sets of 250 hypothetical return samples of different sizes from a normal distribution with the true moments
  - For each simulated set, compute the mean-variance frontier and compare its proximity to the true frontier
  - Frontiers are generated for the unconstrained and the constrained (non-negative weights) case

Limitations of MV portfolios
Jobson and Korkie (1980) experiment

- Findings
  - Mean-variance efficient portfolio weights are very noisy, i.e. they vary a lot even with perfect information on the first two moments
  - Increasing the sample size (i.e. increasing the lookback period) does not substantially reduce the variability
  - Sharpe ratios from plug-in estimates are consistently and considerably lower than those generated from the true frontier
  - Constraints help to reduce the sampling error, but variability still remains high

Jobson and Korkie experiment
Data

10 industry portfolios (monthly returns, Jan 1926 – Dec 2003)

Industry	Mean	Volatility	Correlations									
Non-durables	11.92	16.78	1									
Durables	14.22	25.47	0.76	1								
Manufacturing	12.04	22.25	0.85	0.87	1							
Energy	12.44	20.71	0.64	0.63	0.74	1						
Hightech	13.55	26.08	0.74	0.79	0.86	0.62	1					
Telcom	10.35	16.63	0.65	0.62	0.66	0.5	0.69	1				
Shops	12.13	20.76	0.87	0.8	0.84	0.6	0.79	0.67	1			
Health	13.75	20.37	0.81	0.67	0.77	0.59	0.73	0.59	0.75	1		
Utilities	10.51	20.14	0.71	0.64	0.71	0.62	0.64	0.61	0.66	0.63	1	
Others	11.58	22.59	0.85	0.79	0.91	0.72	0.8	0.69	0.82	0.74	0.75	

Jobson and Korkie experiment
Tangency portfolio

Industry	Unconstrained weights	Constrained weights
Non-durables	81.70%	26.79%
Durables	22.06%	0.00%
Manufacturing	-47.93%	0.00%
Energy	37.07%	21.18%
Hightech	-1.52%	0.00%
Telcom	37.99%	15.61%
Shops	-13.91%	0.00%
Health	28.10%	36.41%
Utilities	-2.54%	0.00%
Others	-41.03%	0.00%
Expected return	12.60%	12.45%
Volatility	14.23%	16.23%
Sharpe ratio	0.6044	0.5165

Jobson and Korkie experiment
Results – Unconstrained efficient frontier

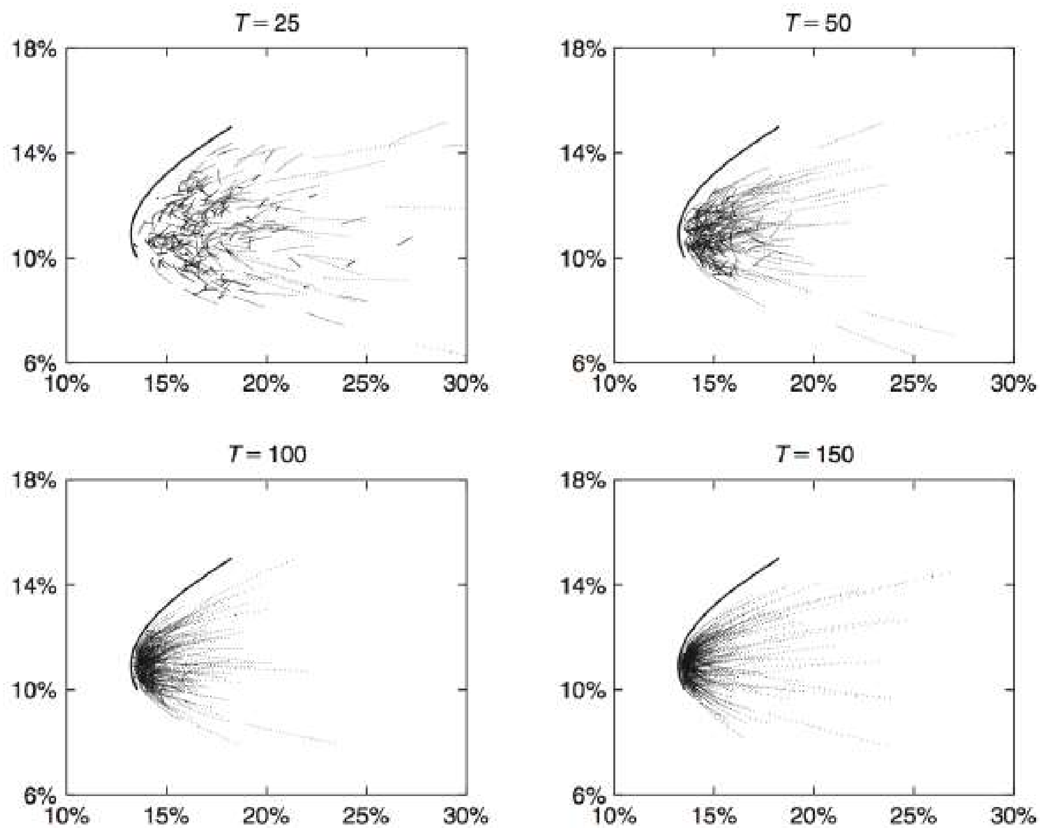


Figure 5.4 The solid line in each plot is the unconstrained mean–variance frontier for 10 industry portfolios, taking sample moments as the truth. The dotted lines show the mean–variance trade-off, evaluated using the true moments, of 250 independent plug-in estimates for 25, 50, 100, and 150 simulated returns.

Source: Brandt M.W. – Portfolio choice problems

Jobson and Korkie experiment
Results – Constrained efficient frontier

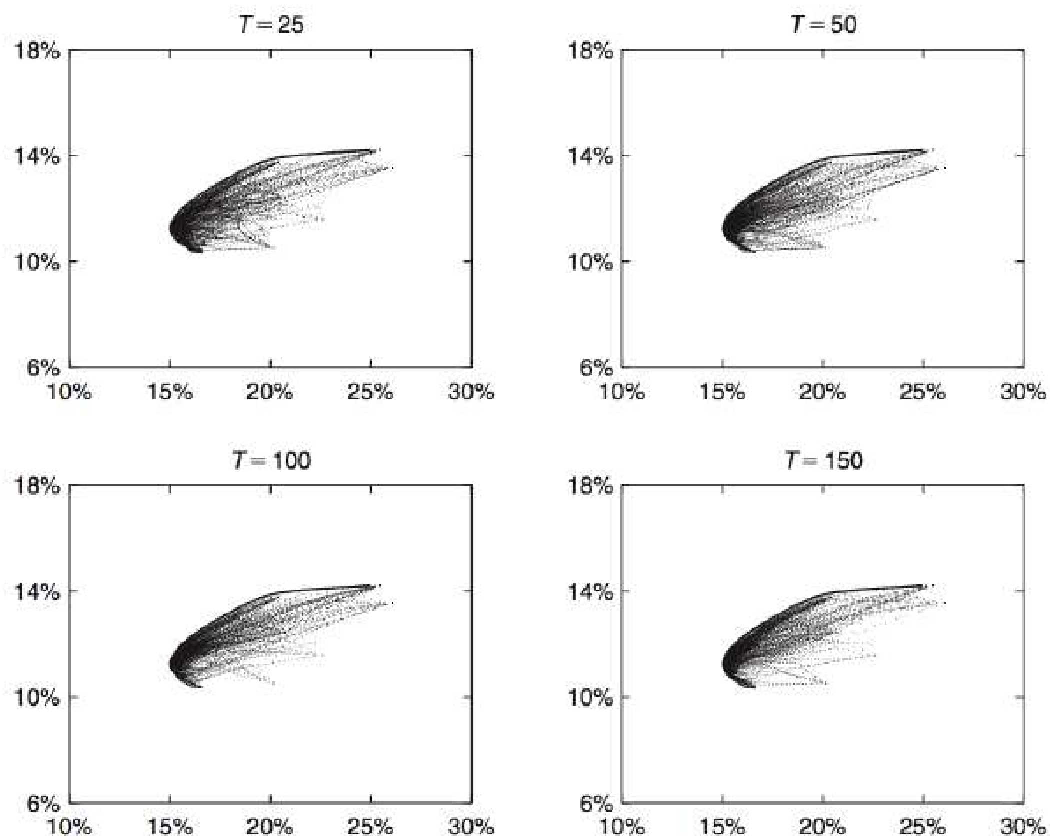


Figure 5.5 The solid line in each plot is the constrained (nonnegative portfolio weights) mean–variance frontier for 10 industry portfolios, taking sample moments as the truth. The dotted lines show the mean–variance trade-off, evaluated using the true moments, of 250 independent plug-in estimates for 25, 50, 100, and 150 simulated returns.

Source: Brandt M.W. – Portfolio choice problems

Jobson and Korkie experiment

Results – Sharpe ratio

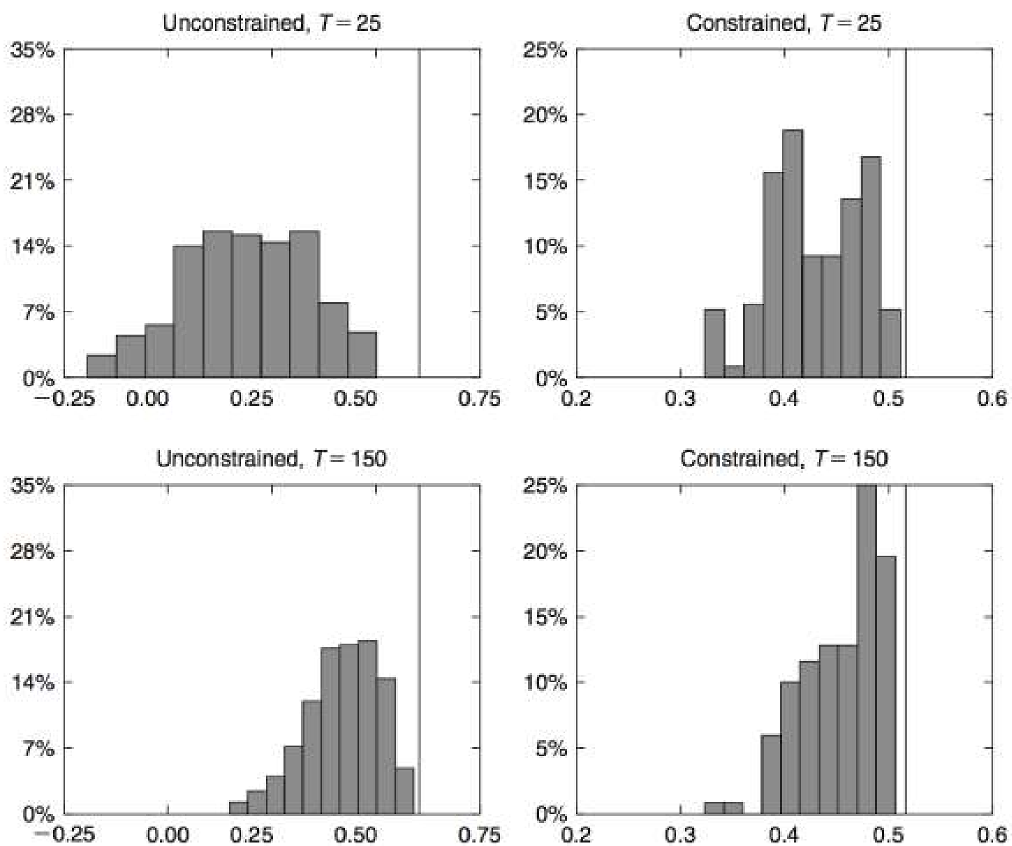


Figure 5.6 The vertical line in each plot represents the Sharpe ratio of the true unconstrained or constrained (nonnegative portfolio weights) tangency portfolios for 10 industry portfolios, taking sample moments as the truth. The histograms correspond to the Sharpe ratios, evaluated using the true moments, of 250 independent plug-in estimates for 25 or 150 simulated returns.

Alternative portfolio construction methods

- **Regularization of optimized portfolios**
  - Introducing portfolio constraints
  - Shrinking the covariance matrix

- **Risk-based weighting schemes**
  - Robust, independent of expected returns
  - Minimum variance, equal risk contribution
  - Volatility targeting

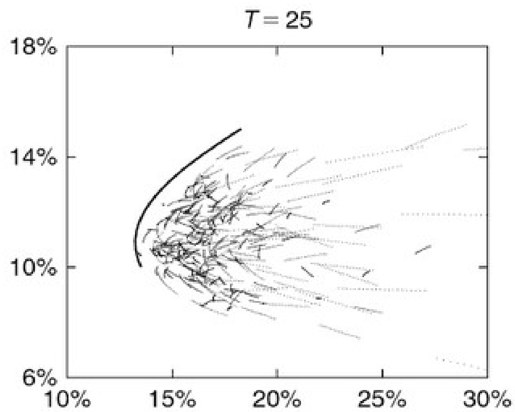
- Black-Litterman approach

- Fundamental law of active management

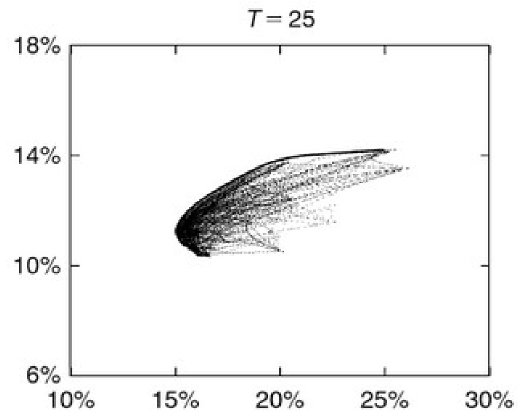
Introducing portfolio constraints

- Effect of no-short sell constraints – Application of Jobson and Korkie (1980) experiment on industry portfolios

Unconstrained



Constrained



- Other constraints include maximum position in stocks, industries or countries. Additional constraints may pertain to controlling for portfolio risk and liquidity characteristics.

- Empirically and in practice, portfolio constraints improve out-of-sample performance of both MV and mean-variance optimal portfolios.

Introducing portfolio constraints

- Green and Hollifield (1992) argue that the presence of a single dominant factor in the covariance matrix of returns is why we observe extreme positive and negative weights. If this were true, then a constrained portfolio should underperform its unconstrained counterpart.

- According to Jagannathan and Ma (2003), imposing portfolio constraints is mathematically analogous to solving an unconstrained problem with shrinkage.

- Since shrinkage reduces the effect of sampling error, imposing portfolio constraints is likely to be beneficial.

Constraints ≡ Covariance shrinkage
Proof

- Consider the following constrained minimum variance (MV) problem for an N-asset portfolio:

$$min_w w^T S w$$

$$s.t. \quad \sum_{i=1}^{N} w_i = 1$$

$$w_i \geq 0, \qquad i = 1,2, \dots, N.$$

$$w_i \leq \bar{w}, \qquad i = 1,2, \dots, N.$$

- The corresponding Lagrange function is then:

$$L(\boldsymbol{w}, \theta, \boldsymbol{\lambda}, \boldsymbol{\delta}) = \boldsymbol{w}^T S \boldsymbol{w} - \theta(\mathbf{1}^T \boldsymbol{w} - 1) - \boldsymbol{\lambda}^T \boldsymbol{w} - \boldsymbol{\delta}^T(\bar{w}\mathbf{1} - \boldsymbol{w})$$

Constraints $\equiv$ Covariance shrinkage
Proof

- The Karush-Kuhn-Tucker (KKT) conditions are then:

$$S\boldsymbol{w} - \theta\mathbf{1} - \boldsymbol{\lambda} + \boldsymbol{\delta} = \mathbf{0} \qquad (1)$$

$$\boldsymbol{\lambda} \geq 0, \text{and } \boldsymbol{\lambda} = 0 \text{ if } \boldsymbol{w} > 0 \qquad \lambda_i w_i = 0 \;\forall\, i \quad (2)$$

$$\boldsymbol{\delta} \geq 0, \text{and } \boldsymbol{\delta} = 0 \text{ if } \boldsymbol{w} < \overline{w} \qquad \delta_i(w_i - \overline{w}) = 0 \;\forall\, i \quad (3)$$

where θ is the Lagrange multiplier related to the linear equality constraint, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$ are the Lagrange multipliers for the non-negativity constraints and $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_N)'$ are the Lagrange multipliers for the upper bound constraint

Constraints $\equiv$ Covariance shrinkage
Proof

- Denote a solution to the earlier constrained portfolio variance minimization problem as $w^{++}(S)$.

- **Proposition 1**: Let
$$\tilde{S} = S + (\boldsymbol{\delta}\mathbf{1}' + \mathbf{1}\boldsymbol{\delta}') - (\boldsymbol{\lambda}\mathbf{1}' + \mathbf{1}\boldsymbol{\lambda}')$$
Then $\tilde{S}$ is symmetric and positive semi-definite, and $w^{++}(S)$ is one of its global minimum variance portfolios.

- Constructing a constrained global MV portfolio from S is equivalent to constructing an unconstrained MV portfolio from $\tilde{S}$.

Constraints ≡ Covariance shrinkage

Proof

- Proof of proposition 1: $\tilde{S}$ is obviously symmetric. To show that it is positive semi-definite, suppose that

$$(w_1, \dots, w_N, \lambda_1, \dots, \lambda_N, \delta_1, \dots, \delta_N, \theta) \equiv (w', \lambda', \delta', \theta)$$

is a solution to the constrained MV portfolio problem.

- For any vector $\boldsymbol{x}$,

$$\begin{aligned} \boldsymbol{x}'\tilde{S}\boldsymbol{x} &= \boldsymbol{x}'S\boldsymbol{x} - \boldsymbol{x}'(\mathbf{1}\boldsymbol{\lambda}' + \boldsymbol{\lambda}\mathbf{1}')\boldsymbol{x} + \boldsymbol{x}'(\mathbf{1}\boldsymbol{\delta}' + \boldsymbol{\delta}\mathbf{1}')\boldsymbol{x} \\ &= \boldsymbol{x}'S\boldsymbol{x} - 2(\boldsymbol{x}'\mathbf{1})\big(\boldsymbol{x}'(\boldsymbol{\lambda} - \boldsymbol{\delta})\big) \qquad (4) \end{aligned}$$

- By the first order condition, $S\boldsymbol{w} - \theta\mathbf{1} = \boldsymbol{\lambda} - \boldsymbol{\delta}$. Hence $\mathbf{x}'(\boldsymbol{\lambda} - \boldsymbol{\delta}) = \boldsymbol{x}'\boldsymbol{S}\boldsymbol{w} - \theta\boldsymbol{x}'\mathbf{1}$. Therefore,

$$2(\boldsymbol{x}'\mathbf{1})\big(\boldsymbol{x}'(\boldsymbol{\lambda} - \boldsymbol{\delta})\big) = 2(\boldsymbol{x}'\mathbf{1})(\boldsymbol{x}'\boldsymbol{S}\boldsymbol{w}) - 2\theta(\boldsymbol{x}'\mathbf{1})^2 \qquad (5)$$

Constraints ≡ Covariance shrinkage
Proof

- Notice that

$$|(\boldsymbol{x}'\mathbf{1})(\boldsymbol{x}'\boldsymbol{S}\boldsymbol{w})| = \left|(\boldsymbol{x}'\mathbf{1})\left(\boldsymbol{x}'S^{1/2}\right)\left(S^{1/2}\boldsymbol{w}\right)\right|$$
$$\leq |(\boldsymbol{x}'\mathbf{1})|(\boldsymbol{x}'S\boldsymbol{x})^{0.5}(\boldsymbol{w}'S\boldsymbol{w})^{0.5}$$

The first equality holds since S is positive semi-definite and the last inequality is due to the Cauchy-Schwarz inequality.

- Using the first order condition again (1)-(3), we get:

$$0 \leq \boldsymbol{w}'S\boldsymbol{w} = \boldsymbol{w}'\boldsymbol{\lambda} - \boldsymbol{w}'\boldsymbol{\delta} + \theta\boldsymbol{w}'\mathbf{1} = \theta - \overline{w}\boldsymbol{\delta}'\mathbf{1} \leq \theta$$

Substituting $\boldsymbol{w}'S\boldsymbol{w}$ in the above ineqaulity, we get:

$$|(\boldsymbol{x}'\mathbf{1})(\boldsymbol{x}'\boldsymbol{S}\boldsymbol{w})| \leq |(\boldsymbol{x}'\mathbf{1})|(\boldsymbol{x}'S\boldsymbol{x})^{0.5}\theta^{0.5} \qquad (6)$$

Constraints ≡ Covariance shrinkage
Proof

- Combining inequality (6) with (4) and (5), we have:

$$\begin{aligned} \boldsymbol{x}'\tilde{S}\boldsymbol{x} &= \boldsymbol{x}'S\boldsymbol{x} - 2(\boldsymbol{x}'\mathbf{1})(\boldsymbol{x}'\boldsymbol{Sw}) + 2\theta(\boldsymbol{x}'\mathbf{1})^2 \\ &\geq \boldsymbol{x}'S\boldsymbol{x} - 2|(\boldsymbol{x}'\mathbf{1})||(\boldsymbol{x}'\boldsymbol{Sw})| + 2\theta(\boldsymbol{x}'\mathbf{1})^2 \\ &\geq \boldsymbol{x}'S\boldsymbol{x} - 2|(\boldsymbol{x}'\mathbf{1})|(\boldsymbol{x}'S\boldsymbol{x})^{0.5}\theta^{0.5} + 2\theta(\boldsymbol{x}'\mathbf{1})^2 \\ &= (a-b)^2 + b^2 \end{aligned}$$

which is strictly non-negative, where $a = (\boldsymbol{x}'S\boldsymbol{x})^{0.5}$ and $b = \theta^{0.5}|(\boldsymbol{x}'\mathbf{1})|$. So $\tilde{S}$ is positive semi-definite.

Constraints ≡ Covariance shrinkage
Proof

- Because $\tilde{S}$ is positive semi-definite, to show that $\boldsymbol{w}$ is an **unconstrained** global minimum variance portfolio of $\tilde{S}$, it suffices to verify the first order condition:

$$
\begin{aligned}
\tilde{S}\boldsymbol{w} &= S\boldsymbol{w} - (\mathbf{1}\boldsymbol{\lambda}' + \boldsymbol{\lambda}\mathbf{1}')w + (\mathbf{1}\boldsymbol{\delta}' + \boldsymbol{\delta}\mathbf{1}')w \\
&= S\boldsymbol{w} - \boldsymbol{\lambda}\mathbf{1}'w + \mathbf{1}\overline{w}(\boldsymbol{\delta}'\mathbf{1}) + \boldsymbol{\delta}\mathbf{1}'\boldsymbol{w} \\
&(\because w_i\lambda_i = 0\ \forall\, i, \delta_i(w_i - \overline{w}) = 0\ \forall\, i) \\
&= S\boldsymbol{w} - \lambda + \overline{w}(\boldsymbol{\delta}'\mathbf{1})\mathbf{1} + \boldsymbol{\delta} \\
&= \theta\mathbf{1} + \overline{w}(\boldsymbol{\delta}'\mathbf{1})\mathbf{1} \\
&= \big(\theta + \overline{w}(\boldsymbol{\delta}'\mathbf{1})\big)\mathbf{1}
\end{aligned}
$$

The fact that $\tilde{S}\boldsymbol{w} = \big(\theta + \overline{w}(\boldsymbol{\delta}'\mathbf{1})\big)\mathbf{1}$ shows that $\boldsymbol{w}$ solves the unconstrained MV portfolio problem for covariance matrix $\tilde{S}$.

Constraints ≡ Covariance shrinkage

Intuition

- Consider the unconstrained *N*-asset global portfolio variance minimization problem. The first order condition is:

$$\sum_{j=1}^{N} w_j S_{i,j} = \lambda_0 \geq 0\,, i = 1, \ldots, N.$$

- Above condition states that at the optimum, stock i's marginal contribution to portfolio variance is the same as stock j's

- Suppose stock i has a higher covariance with other stocks, i.e. the ith row of S has larger elements that other rows, then stock i's marginal contribution will be larger than that of other stocks. To achieve optimality, stock i's weight must be reduced.

Constraints ≡ Covariance shrinkage

Intuition – Non-negativity constraint

- With a nonnegativity constraint, **Proposition 1** becomes

$$\tilde{S} = S - (\mathbf{1}\boldsymbol{\lambda}' + \boldsymbol{\lambda}\mathbf{1}')$$

$$\tilde{S} = S - \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & \lambda_N \end{bmatrix} - \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \end{bmatrix} \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$$

$$= S - \begin{bmatrix} 2\lambda_1 & \cdots & \lambda_1 + \lambda_N \\ \vdots & \ddots & \vdots \\ \lambda_1 + \lambda_N & \cdots & 2\lambda_N \end{bmatrix}$$

- Effect of imposing a constraint is that, whenever a nonnegativity constraint is binding for stock i, its covariances with other stocks, $S_{i,j}, j \neq i$, are reduced by $\lambda_i + \lambda_j > 0$, and its variance is reduced by $2\lambda_i$. Reduction in the ith row of $\tilde{S}$ implies a higher weight for stock i.

Constraints ≡ Covariance shrinkage
Intuition – Upper bound constraints

- With an upper bound constraint, **Proposition 1** becomes
$$\tilde{S} = S + (\boldsymbol{\delta}\mathbf{1}' + \mathbf{1}\boldsymbol{\delta}')$$

- Applying the same principle as before (i.e. non-negativity case), whenever the upper bound is binding for stock i, its variance is raised by $2\delta_i$, and its covariances with another stock j is increased by $\delta_i + \delta_j > 0$.

- Since the largest (smallest) covariance estimates are more likely caused by upward-biased (downward-biased) estimation error in the case of non-negativity constraint (upper bound constraint), shrinkage may reduce the estimation error.

Shrinking the covariance matrix
Characteristics of sample covariance matrix

- Advantages
  - Intuitive and computationally simple
  - Unbiased

- Disadvantages
  - Very imprecise for large number of assets
  - Asset correlation difficult to estimate
  - Poorly conditioned or even singular (not invertible) when number of assets (N) > number of time periods (T)
  - Portfolio optimizer highly sensitive to extreme elements of sample covariance matrix

Shrinking the covariance matrix
Ledoit and Wolf (2003, 2004a, 2004b)

- Consider the sample covariance matrix S and a highly structured estimator, denoted by F. Idea of shrinkage is to linearly combine S and F by $\delta F + (1 - \delta)S$, where $0 < \delta < 1$, the shrinkage intensity.

- The principle of shrinkage is that by properly combining two 'extreme' estimators one can obtain a 'compromise' estimator that performs better than either extreme.

Shrinking the covariance matrix
Ledoit and Wolf (2003, 2004a, 2004b)

- Some candidates for the highly structured estimator, F
  - Identity matrix
  - Single-factor model (Sharpe, 1963)
  - Multi-factor model
  - Constant correlation model
    - Applicable to assets coming from the same asset class, e.g. stocks only
    - All pairwise correlations are identical
    - Estimator of the common constant correlation is simply the average of all sample correlations

- Choice of structured estimator depends on the trade-off between estimation error versus model misspecification

Shrinking the covariance matrix
Formula for shrinkage intensity

- Ledoit and Wolf (2003) propose a quadratic measure of distance between the true and estimated covariance matrices based on the Frobenius norm, defined as:

$$\|Z\|^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} Z_{ij}^2$$

- By considering this distance measure, we arrive at the following quadratic loss function:

$$\mathrm{L}(\delta) = \|\delta F + (1-\delta)S - \Sigma\|^2$$

where Σ is the true covariance matrix.

Shrinking the covariance matrix
Formula for shrinkage intensity

- Goal is to find the shrinkage intensity δ which minimizes the expected loss function:

$$min_{\delta}\mathrm{E}(\|\delta F + (1-\delta)S - \Sigma\|^2)$$

- Under the assumption that *N* (number of assets) is fixed and *T* (time) tends to infinity, Ledoit and Wolf (2003) prove that the optimal δ asymptotically behaves like $\frac{\kappa}{T}$, where κ[1] is a constant.

(1) See Lediot and Wolf, 2003, for details on how κ is obtained

Shrinking the covariance matrix
Empirical study, Lediot and Wolf (2003)

- Optimization problem of a typical active manager:

Minimize:	$x'\Sigma x$	Minimize tracking error or active risk
such that:	$x'\alpha \geq g$	Active return (over the benchmark) is greater than g, some target gain
	$x'\mathbf{1} = 0$	Sum of active weights (i.e. portfolio minus benchmark weights) is zero
	$x \geq -w_B$	No naked short, i.e. not allowed to underweight a position beyond benchmark weight, w_b
	$x \leq c\mathbf{1} - w_B$	Maximum absolute portfolio weight limit of *c*

Shrinking the covariance matrix
Empirical study, Lediot and Wolf (2003)

- Benchmark construction framework
  - US stocks from Feb 1989 to Dec 2002
  - At the beginning of each month, select *N* largest stocks as measured by their market capitalization
  - At the end of each month, we observe the realized stock returns and calculate the value-weighted index.
  - We vary the benchmark size *N,* where *N* = 30, 50, 100, 225, 500.

- To isolate the effect of covariance matrix choice, we fix the ***ex-ante*** information ratio to be equal to 1.5, independently of N, the benchmark size (i.e. we know the realized returns to some degree)

Shrinking the covariance matrix
Empirical study, Lediot and Wolf (2003)

- Alpha forecast construction methodology
  - Let e_{it} denote the excess return of stock i during period t, i.e. the return of stock minus the benchmark return.
  - Let u_{it} be noise terms which are normally distributed with zero mean, and are independent of each other both cross-sectionally and over time.
  - Generate raw forecasts by adding noise to e_{it}, i.e. $raw_{it} = e_{it} + u_{it}$
  - The well-known Fundamental Law of Active Management (FLAM) states that the *ex-ante* information ratio of the manager is determined by the information coefficient (IC) and the breadth of the strategy (number of independent 'bets'): $\mathrm{IR} \approx \mathrm{IC}.\sqrt{\mathrm{Breadth}}$
  - Given that the forecasts are independent of each other and the strategy is rebalanced every month, we have: $\mathrm{Breadth} = 12.N$

Shrinking the covariance matrix
Empirical study, Lediot and Wolf (2003)

- Alpha forecast construction methodology (continued)
  - Fixing the IR at 1.5 and re-arranging the FLAM equation, we get: $\mathrm{IC} = 1.5/\sqrt{12.N}$
  - Smaller *N* yields a higher IC and vice versa (N=30 yields IC=7.9%; N=500 yields IC=1.9%)
  - Normalize raw_{it} to yield scores: $score_{it} = \frac{raw_{it} - \overline{raw_{it}}}{s_{raw_{it}}}$ where $\overline{raw_{it}}$ and $s_{raw_{it}}$ are the mean and standard deviation of the raw forecasts respectively.
  - Finally, scores are transformed into refined forecasts using the relationship:

$$\hat{\alpha}_{it} = s_{e,i}.\mathrm{IC}.score_{it}$$

where $s_{e,i}$ is the volatility of stock i's excess returns
  - Repeat the forecasting process 50 times to smooth out the randomness.

Shrinking the covariance matrix
Empirical study, Lediot and Wolf (2003)

- Evaluation framework
  - At the beginning of each month, feed the following ingredients to the quadratic optimizer.
  - To compute $\hat{\Sigma}$, we use the last T=60 monthly returns of the current benchmark constituents
  - At the end of the month, the realized excess returns is given by $x^T r$ where x is the solution to the quadratic program and r is the vector of stock returns for the month

- Competing covariance matrices
  - No shrinkage target (sample)
  - Single factor model as shrinkage target (Shrink-SF)
  - Constant correlation model as shrinkage target (Shrink-CC)
  - Statistical multi-factor model as shrinkage target (PC-5)

Shrinking the covariance matrix
Empirical study, Lediot and Wolf (2003)

Table 1: **Summary Statistics of Benchmark Returns.** This table presents summary statistics for monthly real returns of several value-weighted benchmark indices. The data range from 02/1983 until 12/2002, yielding 239 returns. The size of the benchmark is denoted by N. All numbers are annualized.

	$N = 30$	$N = 50$	$N = 100$	$N = 225$	$N = 500$
Mean	13.63	13.50	13.29	13.45	13.42
Standard Deviation	15.12	15.02	14.76	14.56	14.52

Source: Lediot and Wolf (2003) – Honey I shrunk the covariance matrix

Shrinking the covariance matrix
Lediot and Wolf (2003) study - Performance

Table 2: **Mean-Summary Statistics for Excess Returns with Gain = 300 bp** This table presents *ex post* information ratios, means, and standard deviations of realized excess returns. The gain (i.e., the expected excess return) was fixed at 300 basis points. The out-of-sample period is 02/1983 until 12/2002, yielding 239 monthly excess returns. The size of the benchmark is denoted by N. 'Sample' denotes the sample covariance matrix; 'Shrink-CC' denotes our shrinkage estimator (2); 'Shrink-SF' denotes the shrinkage estimator of Ledoit and Wolf (2003); 'PC-5' denotes the estimator based on the first five principal components. The results are mean-summaries over 50 repetitions. All numbers are annualized.

	IR	Mean	SD
	$N = 30$		
Sample	0.97	2.18	2.26
Shrink-CC	1.24	2.50	2.03
Shrink-SF	1.18	2.39	2.04
PC-5	1.17	2.45	2.10
	$N = 50$		
Sample	0.79	1.92	2.44
Shrink-CC	1.14	2.21	1.95
Shrink-SF	1.08	2.13	1.98
PC-5	1.11	2.18	1.96

	IR	Mean	SD
	$N = 100$		
Sample	0.59	1.71	2.93
Shrink-CC	0.91	1.87	2.06
Shrink-SF	0.89	1.86	2.10
PC-5	0.91	1.87	2.07
	$N = 225$		
Sample	0.37	2.37	6.45
Shrink-CC	0.54	2.53	4.97
Shrink-SF	0.57	2.37	4.30
PC-5	0.55	2.42	4.46

Source: Lediot and Wolf (2003) – Honey I shrunk the covariance matrix

Shrinking the covariance matrix
Lediot and Wolf (2003) study - Turnover

Table 3: **Mean-Summary Statistics for Average Monthly Turnover** This table presents average monthly turnovers for various strategies. The gain (i.e., the expected excess return) was fixed at 300 basis points. The out-of-sample period is 02/1983 until 12/2002, yielding 239 monthly portfolio updates. The size of the benchmark is denoted by N. 'Sample' denotes the sample covariance matrix; 'Shrink-CC' denotes our shrinkage estimator (2); 'Shrink-SF' denotes the shrinkage estimator of Ledoit and Wolf (2003); 'PC-5' denotes the estimator based on the first five principal components. The results are mean-summaries over 50 repetitions.

	$N = 30$	$N = 50$	$N = 100$	$N = 225$	$N = 500$
Sample	0.39	0.50	0.66	0.80	0.85
Shrink-CC	0.33	0.39	0.50	0.65	0.75
Shrink-SF	0.34	0.41	0.52	0.66	0.76
PC-5	0.33	0.39	0.50	0.64	0.73

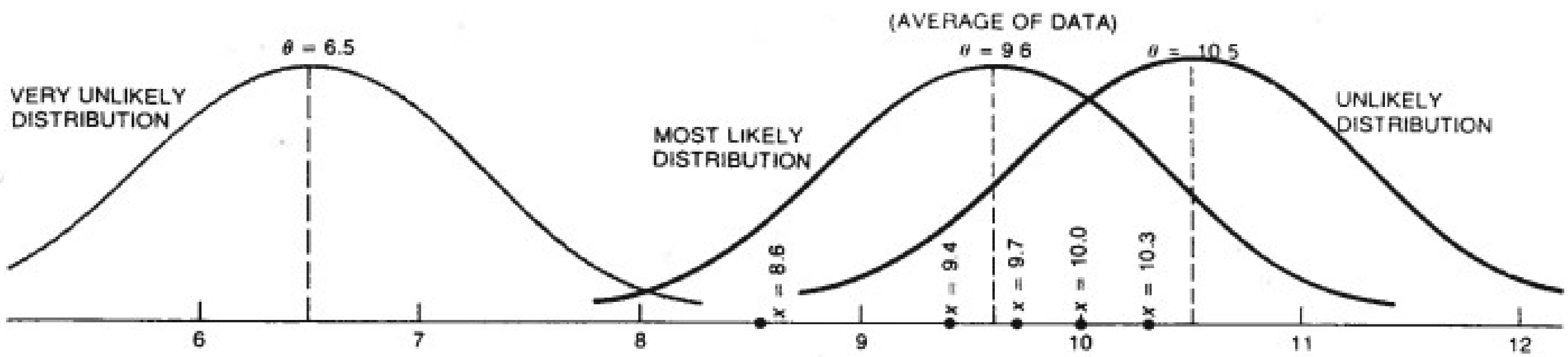
Shrinking the covariance matrix
Key takeaway points

- Shrinkage yields superior information ratios than the sample covariance matrix.
- Ex-post portfolio tracking error is significantly lower when shrinkage is used.
- Portfolio turnover is reduced when the shrinkage technique is employed.

Risk-based weighting schemes
Stein’s paradox

- The best guess about the future is usually obtained by computing the average of past events.
- Stein’s paradox defines circumstances in which there are estimators better than the arithmetic average.
- James Stein shrinkage estimator take the form
$$\hat{\mu}_S(y) = \delta \times \mu_0 + (1 - \delta) \times \hat{\mu} \qquad \text{for} \quad 0 < \delta < 1$$
where $\hat{\mu}_S(y)$ is the shrinkage estimator for the true population y, μ_0 is some fixed value, $\hat{\mu}$ is the vector of sample means and δ is the shrinkage factor.
- Common shrinkage targets: Zero, cross-sectional mean, theoretical values
- Risk-based weighting schemes assume expected returns are equal across all assets, i.e. $\delta = 1$

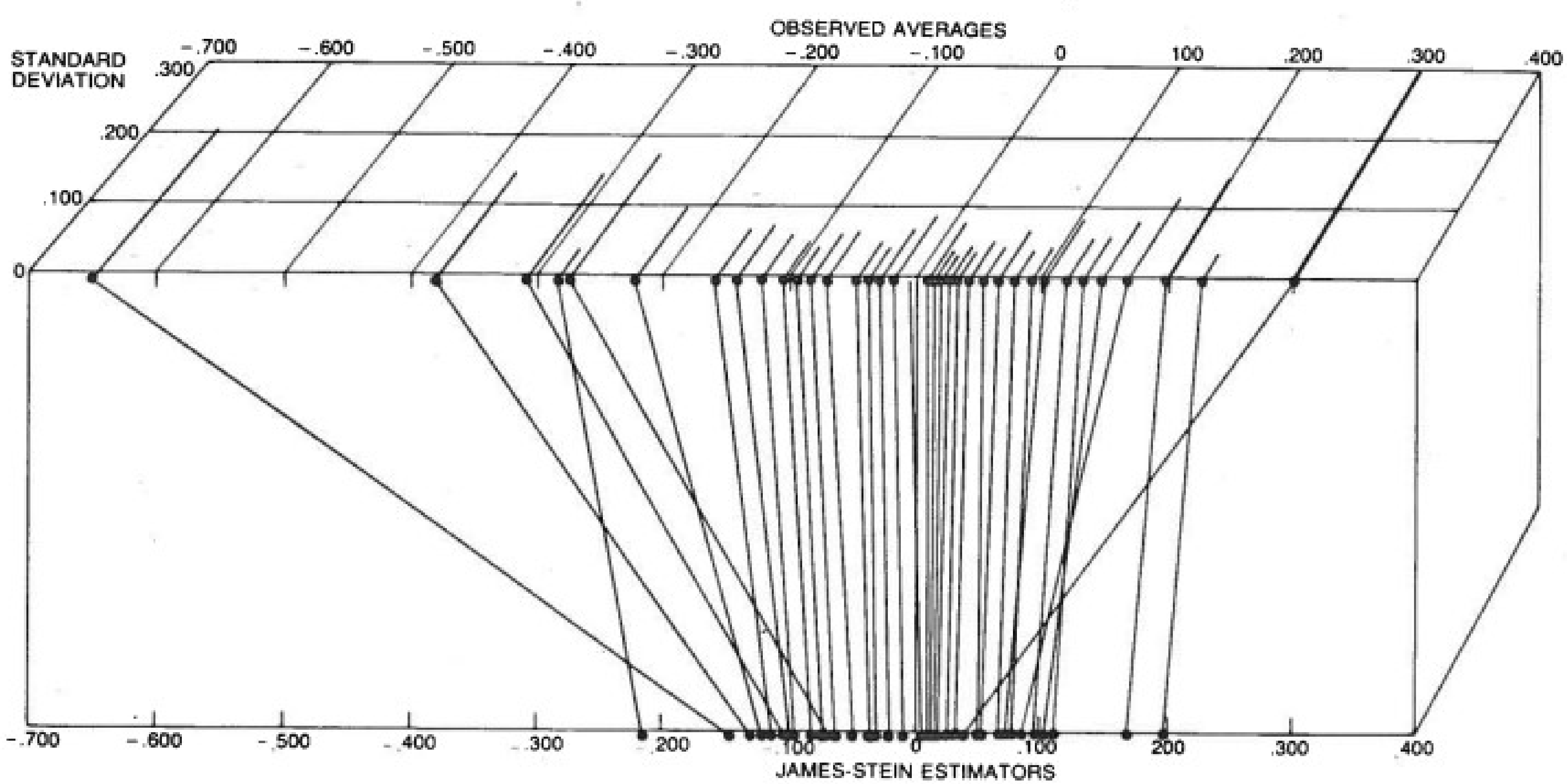
- Paradox: Any shrinkage results in less risk than sample averages



PROBLEM IN STATISTICS is to deduce from a set of data the true mean and standard deviation of the distribution. Even when it is known that the distribution is a normal one and that the standard deviation is 1, the mean could in principle have any value. Some values, however, are more likely than others. For example, the five data points (x) given here could be described by a normal distribution with a mean of 6.5 only if all five points were more than two standard deviations above the mean. It can be shown that the data are most likely to be generated by a distribution with a mean equal to the observed average of the data, denoted $\bar{x}$. In this case the average is equal to 9.6.

Stein’s paradox

An example



SHRINKING of the observed toxoplasmosis rates to yield a set of James-Stein estimators substantially alters the apparent distribution of the disease. The shrinking factor is not the same for all the cities but instead depends on the standard deviation of the rate measured in that city. A large standard deviation implies that a measurement is based on a small sample and is subject to large random fluctuations; that measurement is therefore compressed more than the others are. In the El Salvador data the most extreme observations tend to be correlated with the largest standard deviations, again suggesting the unreliability of those measurements. Compared with the observed rates, the James-Stein estimators can be proved to have a smaller total error of estimation. They also provide a more accurate ranking of the cities.

Alternative weighting schemes

Exhibit 1: Passive and 'Alternative-Passive' Portfolio Construction Approaches

	ASSET WGHTING	REQUIRE COVAR MAT?	REQUIRE REBAL?	ALT. APPROACHES
Capitalization-Weighted	Market Cap Weighted	No	No, except for index changes	Optimized tracking portfolios
1/N Portfolio	Equal Weighted	No	Yes	Sector-Neutral 1/N Portfolio
Fundamental Indexing	Based on fundamental characteristics	No	Yes	Sector-neutrality, Style Blending
Minimum Volatility	Equal Marginal Risk contribution	Yes	Yes	Long-Only constraints, concentration constraints
Equal Risk Contributions	Equal Total Risk contribution	Yes	Yes	
Maximum Entropy	Maximal Cross Entropy	Yes	Yes	

Source: Morgan Stanley QDS (2010) – Weighing weighting schemes risk-conditioned portfolio design

Portfolio risk decomposition

- Consider a portfolio $x = (x_1, x_2, \dots, x_n)$ of n risky assets. Let σ_i^2 be the variance of asset i, σ_{ij} be the covariance between assets i and j, and Σ be the covariance matrix.

- Let $\sigma(x) = \sqrt{x^T \Sigma X} = \sqrt{\sum_i x_i^2 \sigma_i^2 + \sum_i \sum_{j \neq i} x_i x_j \sigma_{ij}}$ be the risk of the portfolio. The marginal risk contribution is defined as:

$$\partial_{x_i} \sigma(x) = \frac{\partial \sigma(x)}{\partial x_i} = \frac{x_i \sigma_i^2 + \sum_{j \neq i} x_j \sigma_{ij}}{\sigma(x)}$$

- If the total risk contribution of the i^{th} asset is $\sigma_i(x) = x_i \times \partial_{x_i} \sigma(x)$, then the portfolio risk can be decomposed as $\sigma(x) = \sum_{i=1}^{n} \sigma_i(x)$

Equal risk contribution (ERC) portfolios

- Specification of the ERC strategy

$$x^* = \left\{ x \in [0,1]^n : \sum x_i = 1 \, , x_i \times \partial_{x_i} \sigma(x) = x_j \times \partial_{x_j} \sigma(x) \; \forall \, i,j \right\}$$

$$= \left\{ x \in [0,1]^n : \sum x_i = 1 \, , x_i \times (\Sigma x)_i = x_j \times (\Sigma x)_j \forall \, i,j \right\}$$

where $(\Sigma x)_i$ denotes the i^{th} row of the vector issued from the product of Σ with x

ERC portfolios

- Recall that the total risk contribution of the i^{th} asset is $\sigma_i(x) = x_i \times \partial_{x_i}\sigma(x) = \frac{x_i^2\sigma_i^2 + \sum_{j\neq i} x_i x_j \rho_{ij}\sigma_i\sigma_j}{\sigma(x)}$

- Equal correlations $\left(\rho_{ij} = \rho \; \forall \; i, j\right)$ and 2-asset case

$$\sigma_i(x) = \left(x_i^2\sigma_i^2 + \rho \sum_{j\neq i} x_i x_j \sigma_i \sigma_j\right) / \sigma(x)$$

$$\sigma_i(x) = x_i\sigma_i\left((1-\rho)x_i\sigma_i + \rho \sum_j x_j\sigma_j\right) \Big/ \sigma(x)$$

$$ERC\ condition: \sigma_i(x) = \sigma_k(x)\ \forall i, k$$

Note that the ERC condition will only hold if $x_i\sigma_i = x_k\sigma_k$

Using the relationship $\sum_{j=1}^{n} x_j = 1$, we get:

$$\sum_{j=1}^{n} \frac{x_i\sigma_i}{\sigma_j} = 1 \quad x_i = \frac{\sigma_i^{-1}}{\sum_j \sigma_j^{-1}} \propto \frac{1}{\sigma_i}$$

- Stocks with higher volatilities have lower weights in ERC portfolios.

ERC portfolios

- Equal volatilities ($\sigma_i = \sigma \; \forall \; i$)

$$\sigma_i(x) = \left(x_i^2\sigma_i^2 + \sigma^2 \sum_{j \neq i} x_i x_j \rho_{ij} \right) / \sigma(x)$$

$$= \left(x_i^2\sigma_i^2 - x_i^2\sigma_i^2 + \sigma^2 \sum_{j} x_i x_j \rho_{ij} \right) / \sigma(x) = \left(x_i \sigma^2 \sum_{j} x_j \rho_{ij} \right) / \sigma(x)$$

$$\text{ERC condition: } \sigma_i(x) = \sigma_k(x) \; \forall i,k \quad x_i \sum_j x_j \rho_{ij} = x_k \sum_j x_j \rho_{kj} \qquad \frac{x_k}{x_i} = \frac{\sum_j x_j \rho_{ij}}{\sum_j x_j \rho_{kj}}$$

$$\text{Summing over k yields: } \sum_k \frac{x_k}{x_i} = \sum_k \frac{\sum_j x_j \rho_{ij}}{\sum_j x_j \rho_{kj}}$$

$$x_i = \frac{\left(\sum_j x_j \rho_{ij}\right)^{-1}}{\sum_k \left(\sum_j x_j \rho_{kj}\right)^{-1}} \propto \frac{1}{\sum_j x_j \rho_{ij}}$$

- In this case, a stock with high correlations with stocks in the portfolio that also have large weights will have a smaller overall weight in the ERC portfolio.

- Note that the solution is endogenous because x_i is a function of itself, both directly and through the constraint $\sum_i x_i = 1$

ERC portfolios
Optimization setup

- Because the solution to the problem is endogenous, it does not have a closed form solution. Finding a solution therefore requires the use of a numerical algorithm, e.g. sequential programming algorithm (SQP)

- Following Maillard et al (2010), we set up the problem as :

$$Min_x \sum_{i=1}^{n} \sum_{j=1}^{n} \left(x_i (\Sigma x)_i - x_j (\Sigma x)_j \right)^2$$

$$\text{u.c. } \mathbf{1}^T x = 1 \text{ and } \mathbf{0} \le x \le \mathbf{1}$$

Summary properties of alternative weighting schemes

- 1/n portfolio: $x_i = x_j$
- MV: $\frac{\partial \sigma(x)}{\partial x_i} = \frac{\partial \sigma(x)}{\partial x_j}$ (equal marginal risk contributions across assets)
- Equal risk contribution: $x_i \times \partial_{x_i}\sigma(x) = x_j \times \partial_{x_j}\sigma(x)$ (equal risk contributions across assets)
- Natural order of portfolio volatilities: $\sigma_{mv} \leq \sigma_{erc} \leq \sigma_{1/n}$
  - Under an alternative optimization setup, Maillard et al (2010) minimize portfolio variance subject to a constraint of sufficient diversification in terms of component weights.
  - Under this framework, they prove the above order of portfolio volatilities, with ERC in between 1/n and minimum variance.

Numerical example
Maillard et al (2010)

- Consider four risky assets with volatilities 10%, 20%, 30% and 40%, with the following correlation matrix

$$\begin{bmatrix} 1 & & & \\ 0.8 & 1 & & \\ 0 & 0 & 1 & \\ 0 & 0 & -0.5 & 1 \end{bmatrix}$$

The solution for the $1/n$ rule is

$\sigma(x) = 11.5\%$	x_i	$\partial_{x_i}\sigma(x)$	$x_i \times \partial_{x_i}\sigma(x)$	$c_i(x)$
1	25%	0.056	0.014	12.3%
2	25%	0.122	0.030	26.4%
3	25%	0.065	0.016	14.1%
4	25%	0.217	0.054	47.2%

The solution for the minimum-variance portfolio is

$\sigma(x) = 8.6\%$	x_i	$\partial_{x_i}\sigma(x)$	$x_i \times \partial_{x_i}\sigma(x)$	$c_i(x)$
1	74.5%	0.086	0.064	74.5%
2	0%	0.138	0.000	0%
3	15.2%	0.086	0.013	15.2%
4	10.3%	0.086	0.009	10.3%

Note: $c_i(x)$ is the risk contribution ratio.

Numerical example
Maillard et al (2010)

- Note that the ERC portfolio is invested in all assets
- In this example, ERC weights are in between MV and 1/N

The solution for the ERC portfolio is

$\sigma(x) = 10.3\%$	x_i	$\partial_{x_i}\sigma(x)$	$x_i \times \partial_{x_i}\sigma(x)$	$c_i(x)$
1	38.4%	0.067	0.026	25%
2	19.2%	0.134	0.026	25%
3	24.3%	0.106	0.026	25%
4	18.2%	0.141	0.026	25%

Note: $c_i(x)$ is the risk contribution ratio.

Empirical evidence
SPX and MSCI Europe

- Historical constituents of MSCI Europe and S&P 500
- Monthly rebalancing frequency
- Historical covariance matrix using 1-year of daily returns. For Europe, the European Barra risk model is also used.
- Model specifications
  - Market cap (MC)
  - Equal weight (EW)
  - MV (minimum variance)
    - MV1 – Europe Barra risk model, max weight of 4%
    - MV2 – Historical covariance matrix, max weight of 4%
    - MV3 – Europe only; additional country and sector constraints, max turnover of 30%
  - ERC (equal risk contribution)

Empirical evidence
SPX and MSCI Europe

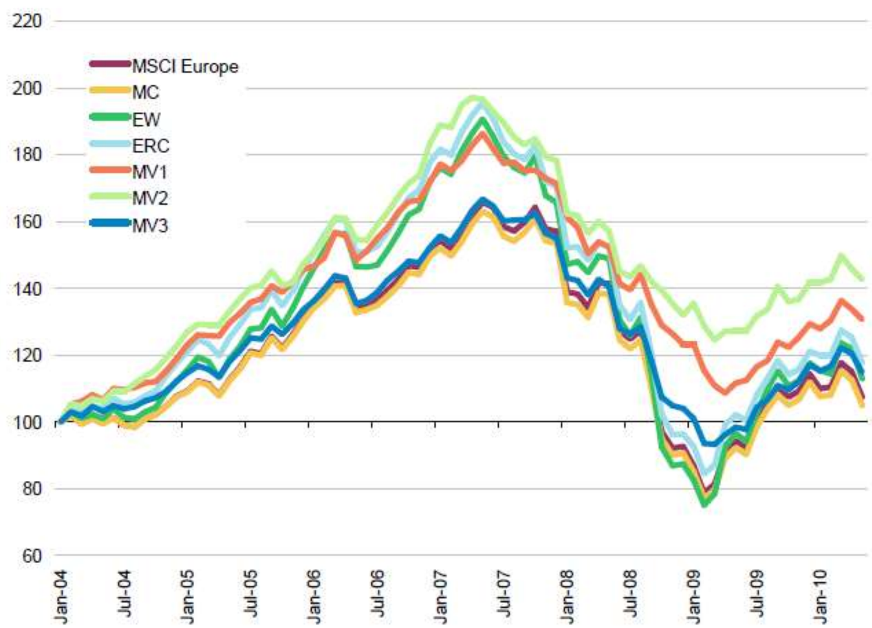
B – MSCI Europe Comparative Statistics (2004 – June 2010)

	MC	EW	ERC	MV1	MV2	MV3
Return (ann)	2.2%	4.3%	4.4%	5.2%	6.5%	3.1%
Volatility (ann)	15.4%	18.5%	15.9%	9.8%	9.9%	11.1%
Monthly VAR (5%)	9.5%	9.6%	9.2%	5.1%	3.5%	7.4%
Turnover (ann)	11.1%	43.5%	70.1%	255.5%	238.6%	139.1%
Maximum Drawdown	52.7%	60.7%	56.8%	41.7%	36.9%	44.1%
Tracking Error (ann)	n/a	4.9%	3.7%	10.5%	10.6%	6.2%
Information Ratio	0.1	0.2	0.3	0.5	0.7	0.3

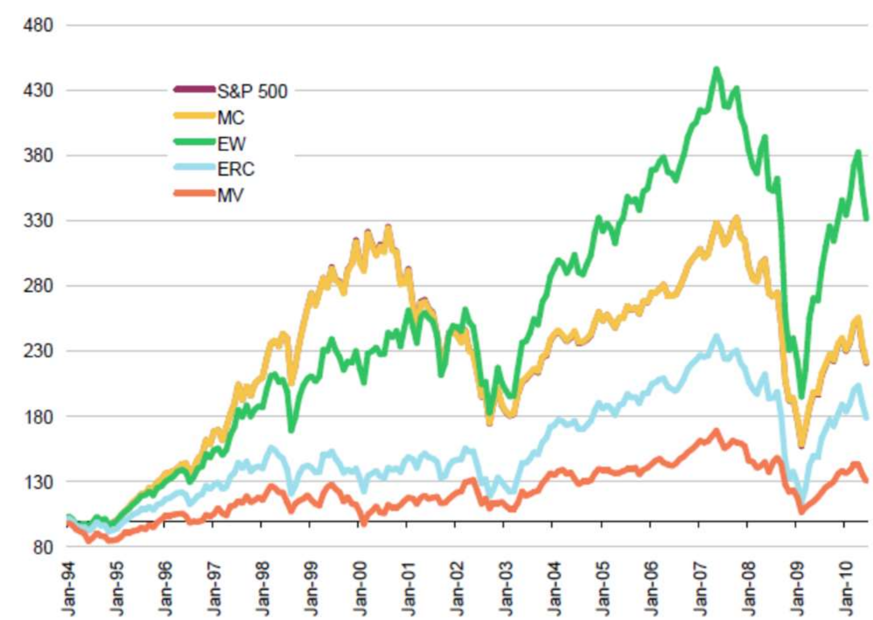
F – S&P 500 Comparative Statistics (1994 – June 2010)

	MC	EW	ERC	MV
Return (ann)	6.1%	8.8%	4.7%	2.3%
Volatility (ann)	15.5%	17.4%	15.1%	11.4%
Monthly VAR (5%)	8.0%	7.7%	7.1%	5.9%
Maximum Drawdown	52.2%	56.2%	52.8%	36.9%
Tracking Error (ann)	n/a	6.2%	7.0%	10.8%
Information Ratio	0.4	0.5	0.3	0.2

A – MSCI Europe Cumulative performance (2004 – June 2010)



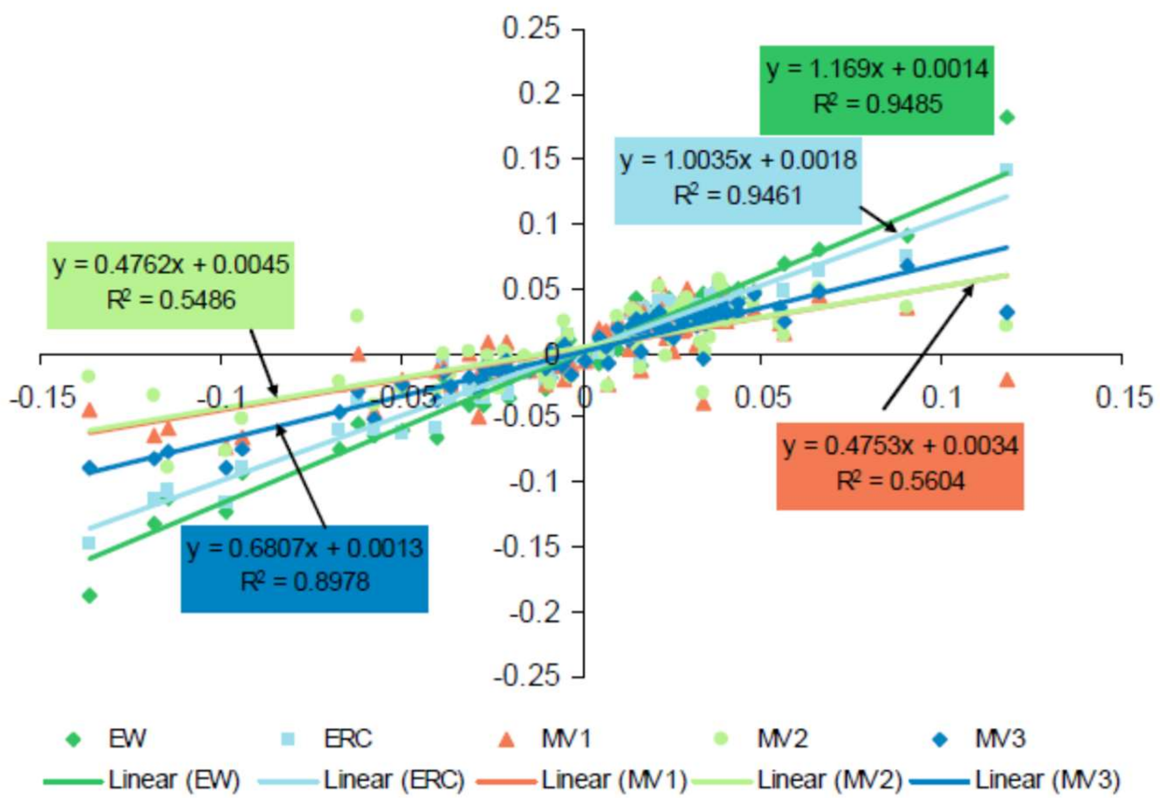
E – S&P 500 Cumulative performance (1994 – June 2010)



Source: Morgan Stanley Quantitative & Derivatives Strategies (2010) – Weighing weighting schemes

Empirical evidence
Beta profile

Exhibit 3: MSCI Europe market beta

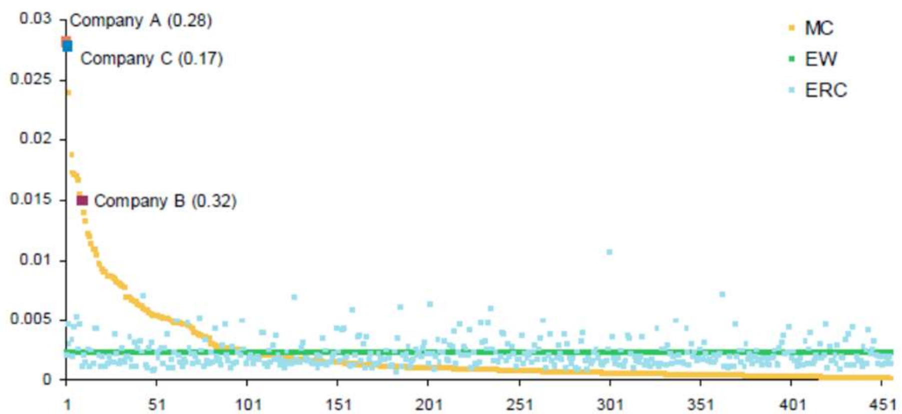


Source: Morgan Stanley Quantitative & Derivatives Strategies (2010) – Weighing weighting schemes

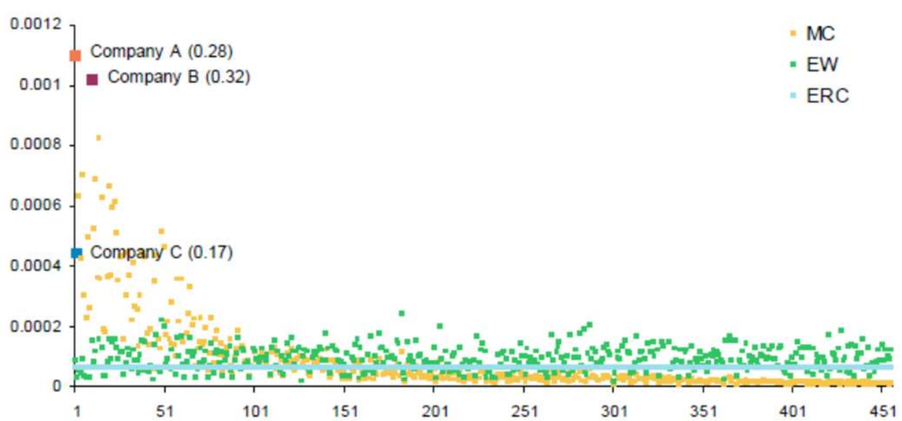
Empirical evidence
Stock weights and risk contribution

Exhibit 4: MSCI Europe Cross-sectional Portoflio Weights - May 10

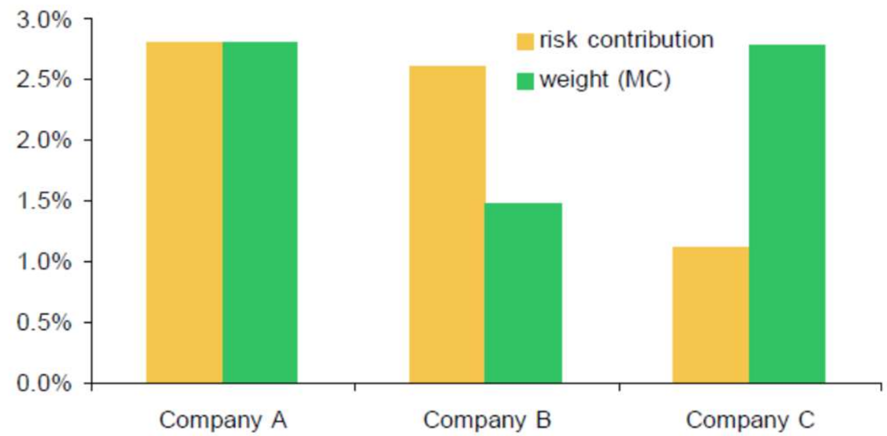
A – Weights



B – Total Risk Contributions



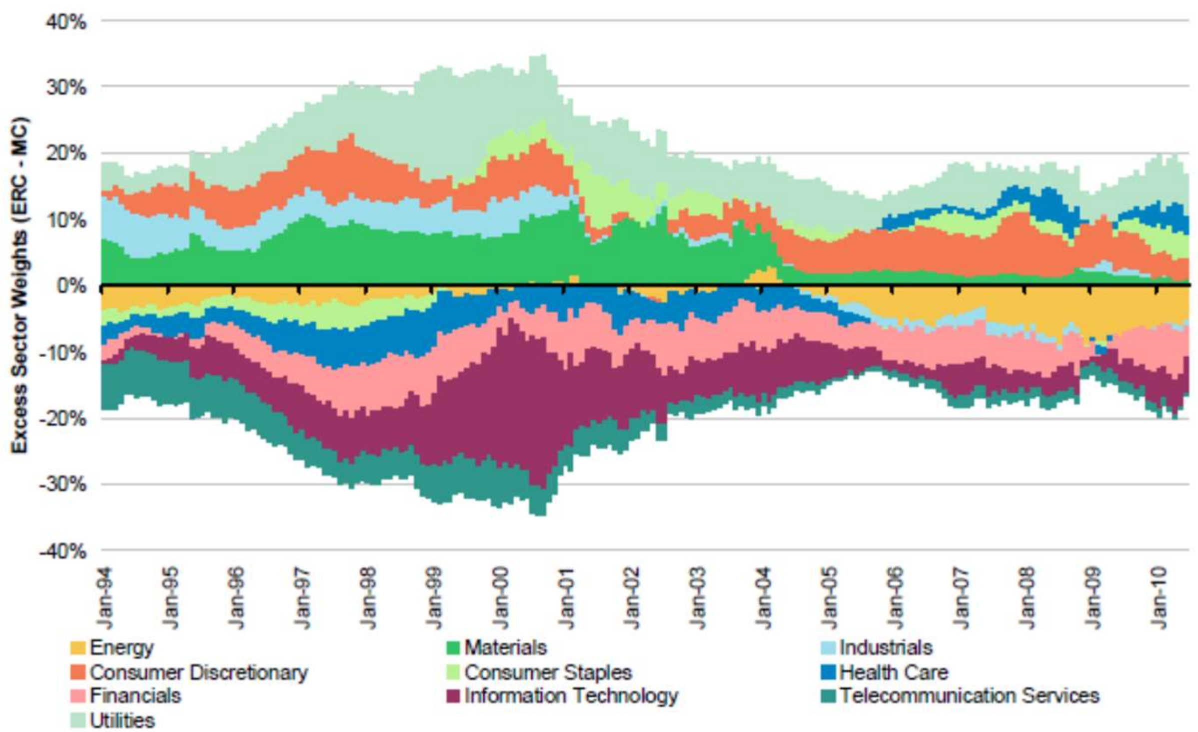
C – Top Risk Contributions and Weights



Source: Morgan Stanley Quantitative & Derivatives Strategies (2010) – Weighing weighting schemes

Empirical evidence
Sector tilts relative to market cap weighting

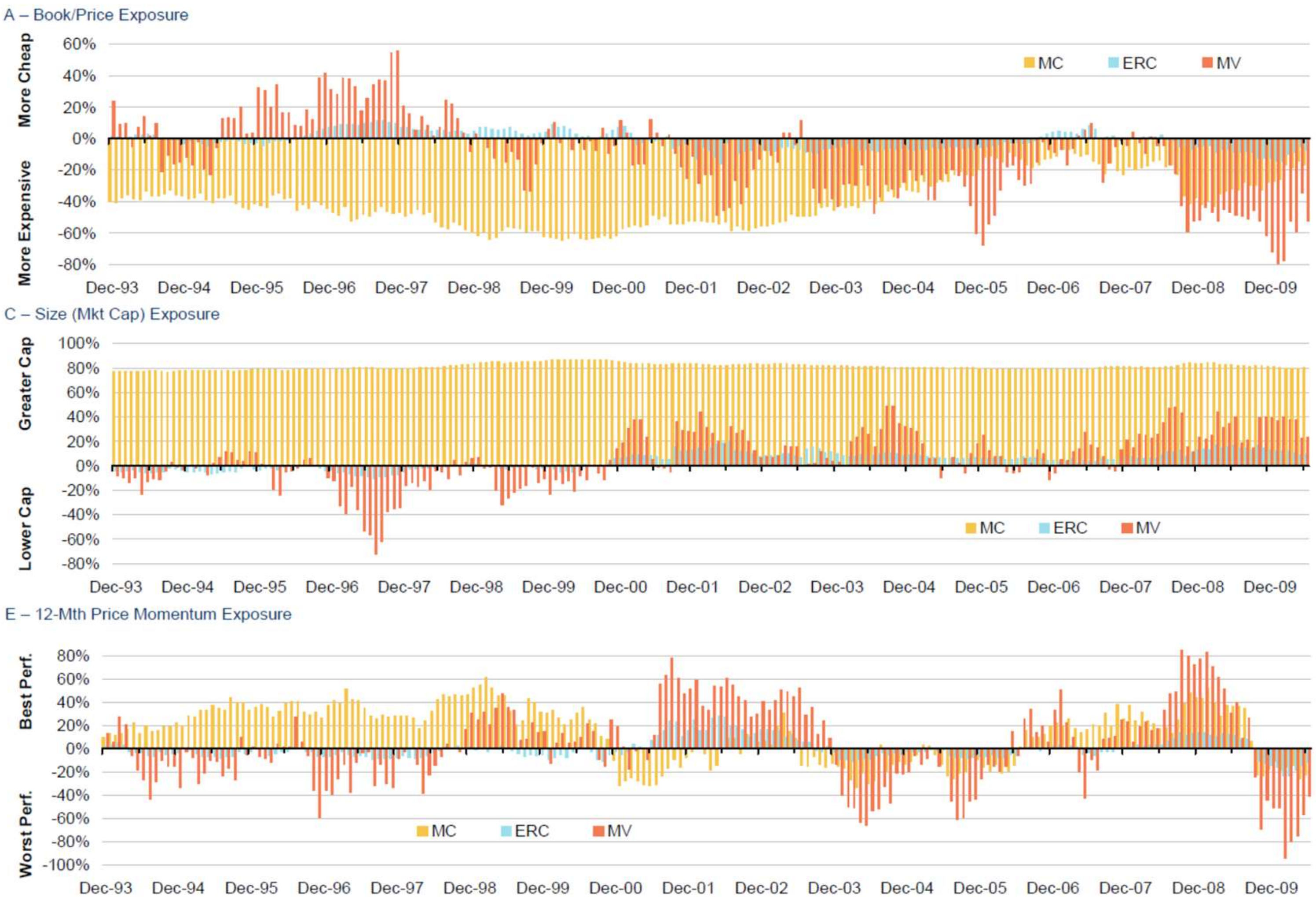
B – S&P 500



Source: Morgan Stanley Quantitative & Derivatives Strategies (2010) – Weighing weighting schemes

Empirical evidence
Style tilts

Exhibit 10: S&P 500 Style Allocations



Source: Morgan Stanley Quantitative & Derivatives Strategies (2010) – Weighing weighting schemes

Volatility targeting strategies

- Volatility targeting strategies adjust the exposure to some asset or portfolio according to an estimate of historical volatility and hence exhibit ex-post an approximately flat volatility profile over time.

- Over (under) weight asset/ portfolio during low (high) volatility states

$$r_{t+1}^{Vol\,managed} = \frac{\sigma_{Target}}{\sigma_t} r_{t+1}$$

where $\hat{\sigma}_t$ is the volatility estimate and σ_{Target} is the volatility target.

Stock picking under different regimes

Eureka Asia Ex Japan Long-Short Equity Index

VIX Level	Dispersion: Low	Dispersion: High
Low	1.1%	3.7%
Medium	1.0%	2.0%
High	-1.2%	-0.1%

Eureka Japan Long-Short Equity Index

VIX Level	Dispersion: Low	Dispersion: High
Low	0.1%	3.0%
Medium	0.8%	0.8%
High	-0.8%	-0.1%

Eureka Europe Long-Short Equity Index

VIX Level	Dispersion: Low	Dispersion: High
Low	1.9%	4.2%
Medium	1.9%	2.0%
High	0.6%	0.2%

Eureka North America Long-Short Equity Index

VIX Level	Dispersion: Low	Dispersion: High
Low	0.6%	1.8%
Medium	1.0%	1.0%
High	-0.6%	-0.7%

Best **Worst**

Source: Morgan Stanley Quantitative & Derivatives Strategies

Volatility targeting in practice
Managing momentum risk

- Barroso and Santa-Clara (2012) find that momentum volatility is highly predictable, even more so than market volatility which is well-known to be predictable.

- Scaling the momentum portfolio to achieve constant ex-ante volatility almost doubles the Sharpe ratio of the vanilla momentum strategy. The authors find the Sharpe ratio improvement robust across subsamples and other major markets.

- Decomposing the volatility of momentum into a market and specific component, the authors find that
  - Most of the risk of momentum is specific to the strategy
  - Specific risk is more predictable than the market component

Predictability of factor volatility

AR(1)specification: $RV_{i,t} = \alpha + \rho RV_{i,t-1} + \varepsilon_t$

Table 2

AR(1) of one-month realized variances

The realized variances are the sum of squared daily returns in each month. The AR(1) regresses the non-overlapping realized variance of each month on its own lagged value and a constant. The out-of-sample (OOS) R-square uses the first 240 months to run an initial regression, so producing an OOS forecast. It then uses an expanding window of observations until the end of the sample. In Panel A, the sample period is from 1927:03 to 2011:12. In Panel B, we repeat the regressions for RMRF (market risk factor) and WML (winner minus losers) and add the same information for HML (high minus low) and SMB (small minus big). The last two columns show, respectively, the average realized volatility and its standard deviation.

Portfolio	α (t-statistic)	ρ (t-statistic)	R^2	OOS R^2	$\overline{\sigma}$	σ_σ
Panel A: 1927:03 to 2011:12						
RMRF	0.0010	0.60	36.03	38.81	14.34	9.97
	(6.86)	(23.92)				
WML	0.0012	0.70	49.10	57.82	17.29	13.64
	(5.21)	(31.31)				
Panel B: 1963:07 to 2011:12						
RMRF	0.0009	0.58	33.55	25.46	13.76	8.48
	(5.65)	(17.10)				
SMB	0.0004	0.33	10.68	-8.41	7.36	3.87
	(8.01)	(8.32)				
HML	0.0001	0.73	53.55	53.37	6.68	4.29
	(4.88)	(25.84)				
WML	0.0009	0.77	59.71	55.26	16.40	13.77
	(3.00)	(29.29)				

Source: Barroso, P. and Santa-Clara (2012) – Momentum has its moments

Decomposition of WML risk

- Decomposition of WML risk into market and specific risk:

$$RV_{WML,t} = \beta_t^2 RV_{rmrf,t} + \sigma_{e,t}^2$$

Table 4
Decomposition of the risk of momentum

Each row shows the results of an AR(1) for six-month, non-overlapping periods. The first row is for the realized variance of the WML (winners minus losers); the second one, the realized variance of the market. The third row is squared beta, estimated as a simple regression of 126 daily returns of the WML on RMRF (market risk factor). The fourth row is the systematic component of momentum risk; the last row, the specific component. The out of sample R-squares use an expanding window of observations after an initial in-sample period of 20 years.

Variable	α (t-statistic)	ρ (t-statistic)	R^2	R^2_{OOS}
σ^2_{wml}	0.0012	0.70	48.67	43.82
	(2.59)	(12.58)		
σ^2_{rmrf}	0.0012	0.50	24.53	6.70
	(4.29)	(7.37)		
β^2	0.3544	0.21	4.59	5.33
	(6.05)	(2.83)		
$\beta^2\sigma^2_{rmrf}$	0.0007	0.47	21.67	20.87
	(2.73)	(6.80)		
σ^2_{ε}	0.0007	0.72	52.21	47.06
	(2.69)	(13.51)		

Source: Barroso, P. and Santa-Clara (2012) – Momentum has its moments

Risk-managed momentum
International evidence

US

Table 3
Momentum and the economic gains from scaling

The first row presents as a benchmark the economic performance of plain momentum (WML) from 1927:03 to 2011:12. The second row presents the performance of risk-managed momentum (WML*). The risk-managed momentum uses the realized variance in the previous six months to scale the exposure to momentum. The mean, the standard deviation, the Sharpe ratio, and the information ratio are annualized. To obtain an information ratio that does not depend on the volatility target we divided previously both (WML) and (WML*) by their respective standard deviations.

Portfolio	Maximum	Minimum	Mean	Standard deviation	Kurtosis	Skewness	Sharpe ratio
WML	26.18	-78.96	14.46	27.53	18.24	-2.47	0.53
WML^*	21.95	-28.40	16.50	16.95	2.68	-0.42	0.97

International

Table 5
The international evidence

The performance of plain momentum (WML) and scaled momentum (WML*) in the major non-US markets. Original data are from Datastream. The returns for country-specific momentum portfolios are from Chaves (2012). The scaled momentum uses the realized volatility in the previous six months, obtained from daily data. The information ratio of the scaled momentum takes plain momentum as the benchmark. The mean return, the standard deviation, the Sharpe ratio, and the information ratio are all annualized. To obtain an information ratio that does not depend on the volatility target we divided previously both WML and WML* by their respective standard deviations. The returns are from 1980:07 to 2011:10.

Statistic	France		Germany		Japan		UK	
	WML	WML*	WML	WML*	WML	WML*	WML	WML*
Maximum	35.43	31.88	22.04	19.34	17.21	16.61	23.23	45.98
Minimum	-28.11	-15.50	-22.56	-11.11	-29.51	-21.56	-36.47	-22.23
Mean	13.19	17.11	18.37	21.02	1.65	4.21	18.86	40.46
Standard deviation	19.66	16.57	18.10	15.08	19.64	17.88	17.24	22.81
Kurtosis	6.85	5.17	4.02	1.45	4.01	1.96	11.78	4.95
Skewnewss	-0.16	0.73	-0.05	0.33	-0.83	-0.62	-1.43	0.56
Sharpe ratio	0.67	1.03	1.02	1.39	0.08	0.24	1.09	1.77
Information ratio	—	0.90	—	0.73	—	0.34	—	1.19

Source: Barroso, P. and Santa-Clara (2012) – Momentum has its moments