

**SFMW**

*October 15<sup>th</sup> , 2008*



# Convex Filtering of Covariance Matrices

**François Oustry** [francois.oustry@raisepartner.com](mailto:francois.oustry@raisepartner.com)  
**Véronique Piolle** [veronique.piolle@raisepartner.com](mailto:veronique.piolle@raisepartner.com)

<http://www.raisepartner.com>



## Outline

### Introduction to the Semidefinite world

- Semidefinite matrices: few definitions and theorems
- The set of positive semidefinite matrices

### Why filtering matrices?

- Random matrices
- Data issues
- Application in portfolio optimization – robustness and stability

### Frequently used practices

- Increasing the smallest eigenvalues
- Factorial models
- Shrinkage

### Going further

- Optimization-driven filtering
- Semidefinite programming formulation of filtering
- Algorithm

### Example: impact of filtering on investment strategies

## Introduction to the SDP world: some definitions and results

- **Characterization of PSD matrices**

A real symmetric matrix is PSD  
 iff all its eigenvalues are non negative

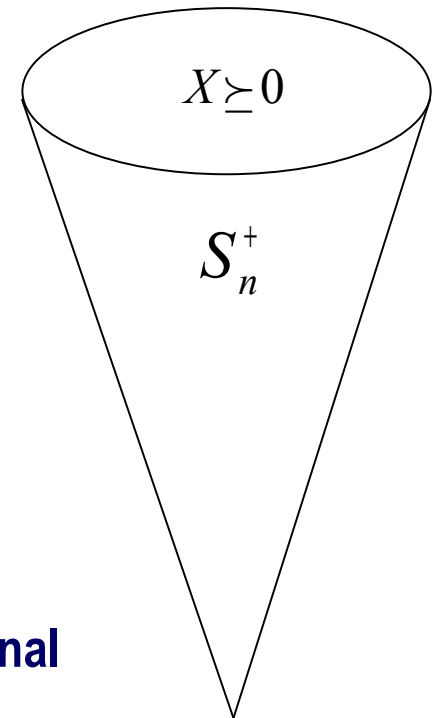
- **Condition number (must be small)**

For PSD matrices:  $\frac{\lambda_{\max}}{\lambda_{\min}}$

- **The set of covariance matrices is exactly the set of PSD matrices**

- **The set of correlation matrices is exactly the set of PSD matrices with unit diagonal**

The cone of PSD matrices

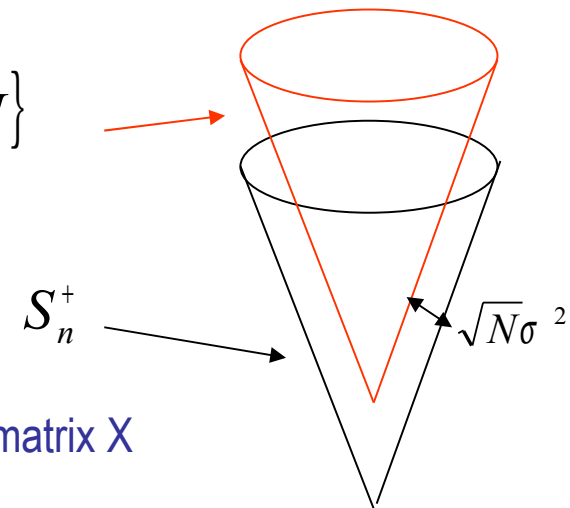


## Introduction to the SDP world: some important results

**Lemma:**  $X \succeq \sigma^2 I$  or  $X - \sigma^2 I \succeq 0 \Leftrightarrow \forall \lambda \in \text{spec}(X), \lambda \geq \sigma^2$

$$\text{Spec}(X - \sigma^2 I) = \text{Spec}(X) - \sigma^2 \mathbf{1}$$

$\{X \text{ symmetric such as } X \succeq \sigma^2 I\}$



### Theorem: Von Neumann lemma

If  $(\lambda^i(X))_{1 \leq i \leq n}$  is the sorted spectrum of matrix  $X$   
(even if  $A$  and  $B$  don't commute)

$$\forall (A, B) \in S_n \times S_n, A - B \succeq 0 \Rightarrow \forall i \in \{1, \dots, n\}, \lambda_A^i \geq \lambda_B^i$$

notation :  $\Lambda(A) \geq \Lambda(B)$

## Why should we filter matrices?

### Historical covariance matrices can be dangerous

#### → Filtering noise / focus on the smallest eigenvalues

Statistical analysis shows the structural deficit of the empirical covariance matrix estimator to capture properly:

- the small contributions to risk
- i.e the small eigenvalues of the covariance matrix
- Or Principal Component variances in the PCA framework

→ highly dangerous for the design of optimal control decisions (optimization or simulation)

→ typical situation when using noisy data

## Why filtering matrices?

### Random matrix theory

The empirical correlation matrix  $\mathbf{C}$  is constructed from the time series of price changes<sup>a</sup>  $\delta x_i(t)$  (where  $i$  labels the asset and  $t$  the time) through the equation:

$$\mathbf{C}_{ij} = \frac{1}{T} \sum_{t=1}^T \delta x_i(t) \delta x_j(t). \quad (0.1)$$

We can symbolically write Eq. (0.1) as  $\mathbf{C} = 1/T \mathbf{M} \mathbf{M}^T$ , where  $\mathbf{M}$  is a  $N \times T$  rectangular matrix, and  $^T$  denotes matrix transposition. The null hypothesis of independent assets, which we consider now, translates itself in the assumption that the coefficients  $M_{it} = \delta x_i(t)$  are independent, identically distributed, random variables<sup>b</sup>; the so-called random Wishart matrices or Laguerre ensemble of the Random Matrix theory<sup>8,10</sup>. We will note  $\rho_C(\lambda)$  the density of eigenvalues of  $\mathbf{C}$ ,

## Why filtering matrices?

### Random matrix theory

📖 L. Laloux, P. Cizeau, J-P. Bouchaud and M. Potters, *Noise Dressing of Financial Correlation Matrices*. Phys. Rev. Lett. 83, 1467 - 1470 , 1999.

Random Matrix theory<sup>8,10</sup>. We will note  $\rho_C(\lambda)$  the density of eigenvalues of  $\mathbf{C}$ , defined as:

**Density of eigenvalues for random matrices = full noise matrices**

$$\rho_C(\lambda) = \frac{1}{N} \frac{dn(\lambda)}{d\lambda}, \quad (0.2)$$

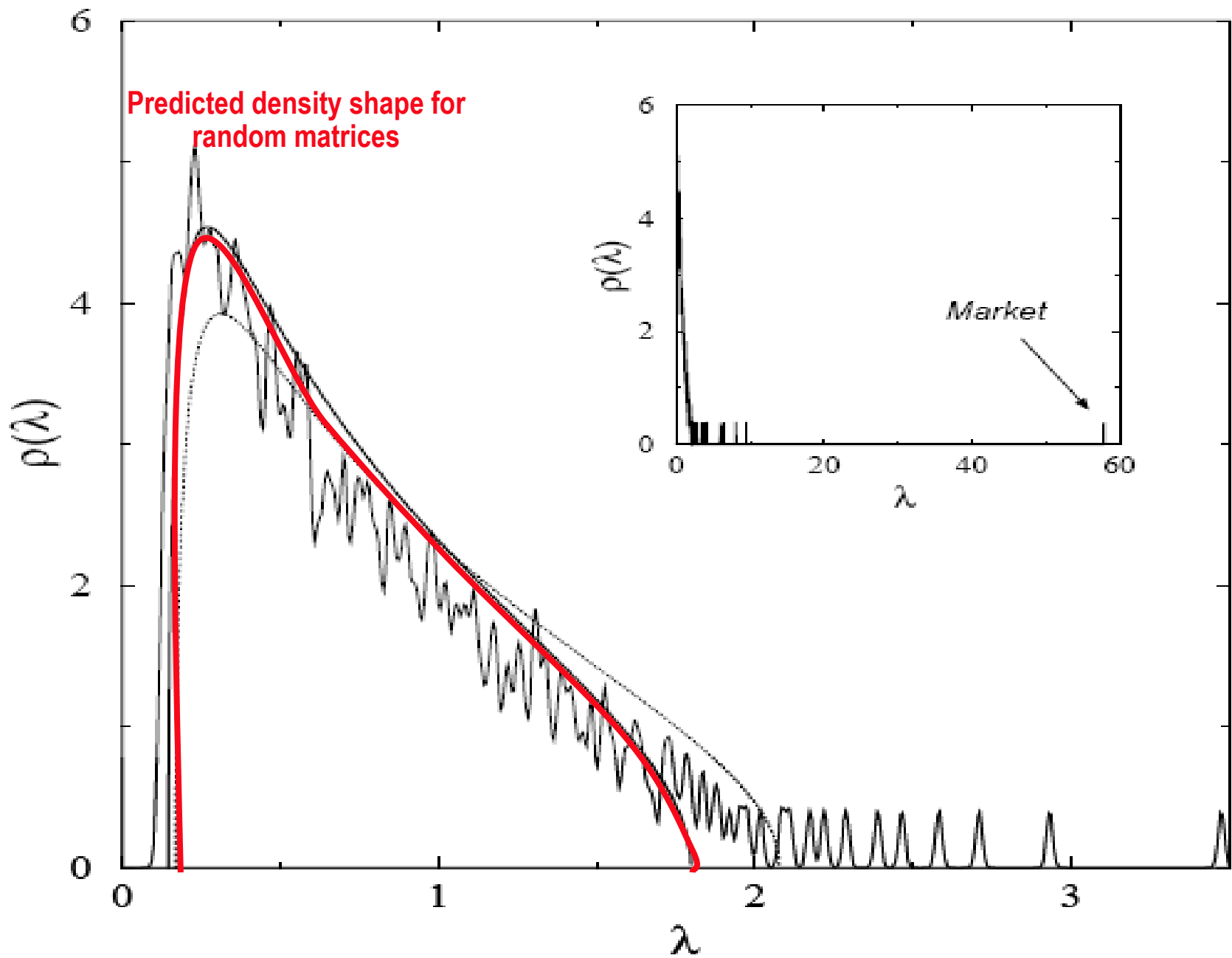
where  $n(\lambda)$  is the number of eigenvalues of  $\mathbf{C}$  less than  $\lambda$ .

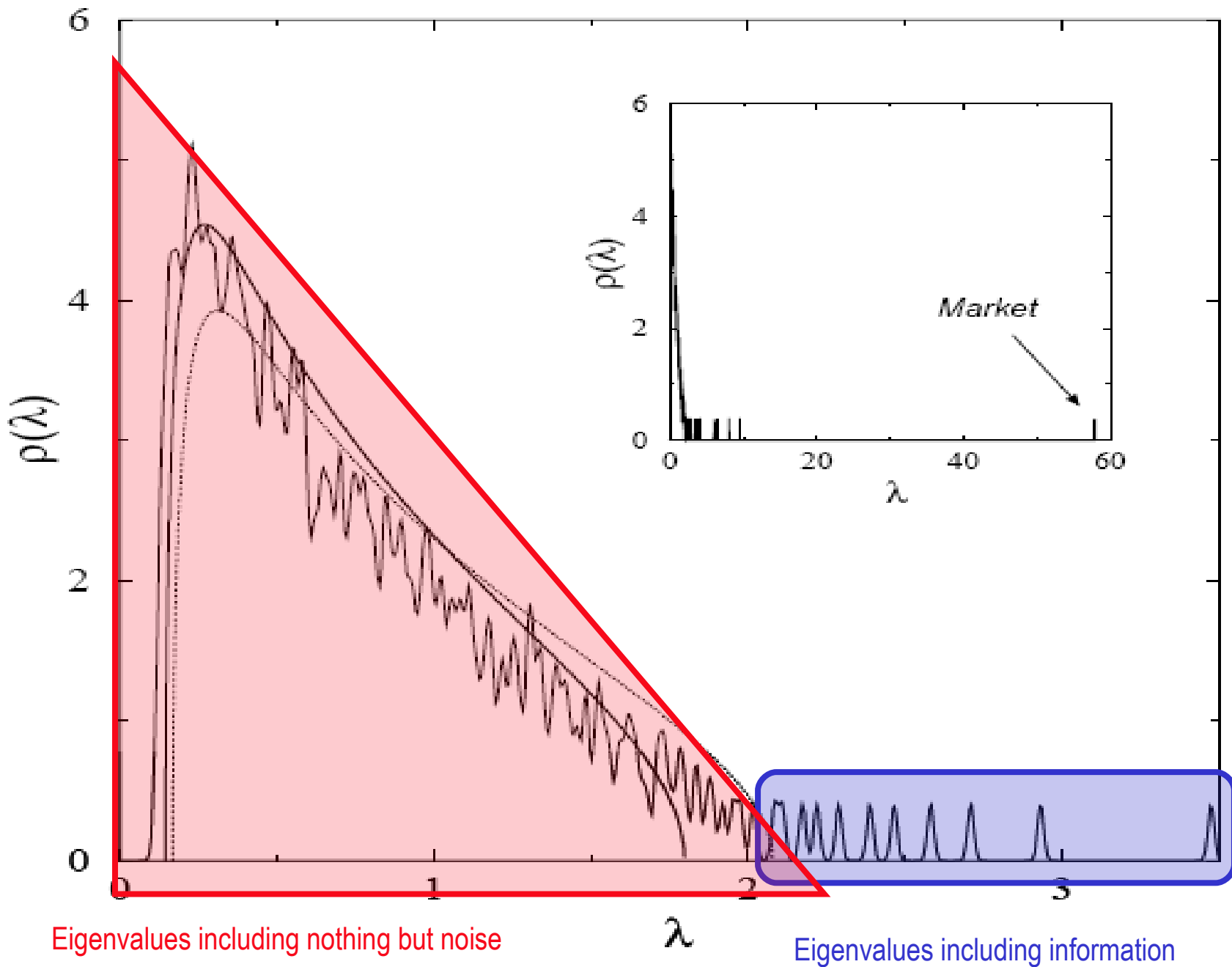
Interestingly, if  $\mathbf{M}$  is a  $T \times N$  random matrix,  $\rho_C(\lambda)$  is self-averaging and exactly known in the limit  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  and  $Q = T/N \geq 1$  fixed<sup>8,9</sup>, and reads:

$$\begin{aligned} \rho_C(\lambda) &= \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \\ \lambda_{min}^{max} &= \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}), \end{aligned} \quad (0.3)$$

with  $\lambda \in [\lambda_{min}, \lambda_{max}]$ , and where  $\sigma^2$  is equal to the variance of the elements of  $\mathbf{M}$ <sup>9</sup>, equal to 1 with our normalisation. In the limit  $Q = 1$  the normalised eigenvalue

Predicted density shape for random matrices





## Why filtering matrices?

### Random matrix theory

 L. Laloux, P. Cizeau, J-P. Bouchaud and M. Potters, *Noise Dressing of Financial Correlation Matrices*. Phys. Rev. Lett. 83, 1467 - 1470 , 1999.

Random Matrix theory<sup>8,10</sup>. We will note  $\rho_C(\lambda)$  the density of eigenvalues of  $\mathbf{C}$ , defined as:

$$\rho_C(\lambda) = \frac{1}{N} \frac{dn(\lambda)}{d\lambda}, \quad (0.2)$$

where  $n(\lambda)$  is the number of eigenvalues of  $\mathbf{C}$  less than  $\lambda$ .

Interestingly, if  $\mathbf{M}$  is a  $T \times N$  random matrix,  $\rho_C(\lambda)$  is self-averaging and exactly known in the limit  $N \rightarrow \infty$ ,  $T \rightarrow \infty$  and  $Q = T/N \geq 1$  fixed<sup>8,9</sup>, and reads:

The quality of the estimation of the smallest eigenvalue is bad when T is small

$$\begin{aligned} \rho(\lambda) &= \frac{Q}{2\pi\sigma^2} \frac{\sqrt{(\lambda_{max} - \lambda)(\lambda - \lambda_{min})}}{\lambda}, \\ \lambda_{min}^{max} &= \sigma^2(1 + 1/Q \pm 2\sqrt{1/Q}), \end{aligned} \quad (0.3)$$

with  $\lambda \in [\lambda_{min}, \lambda_{max}]$ , and where  $\sigma^2$  is equal to the variance of the elements of  $\mathbf{M}$ <sup>9</sup>, equal to 1 with our normalisation. In the limit  $Q = 1$  the normalised eigenvalue

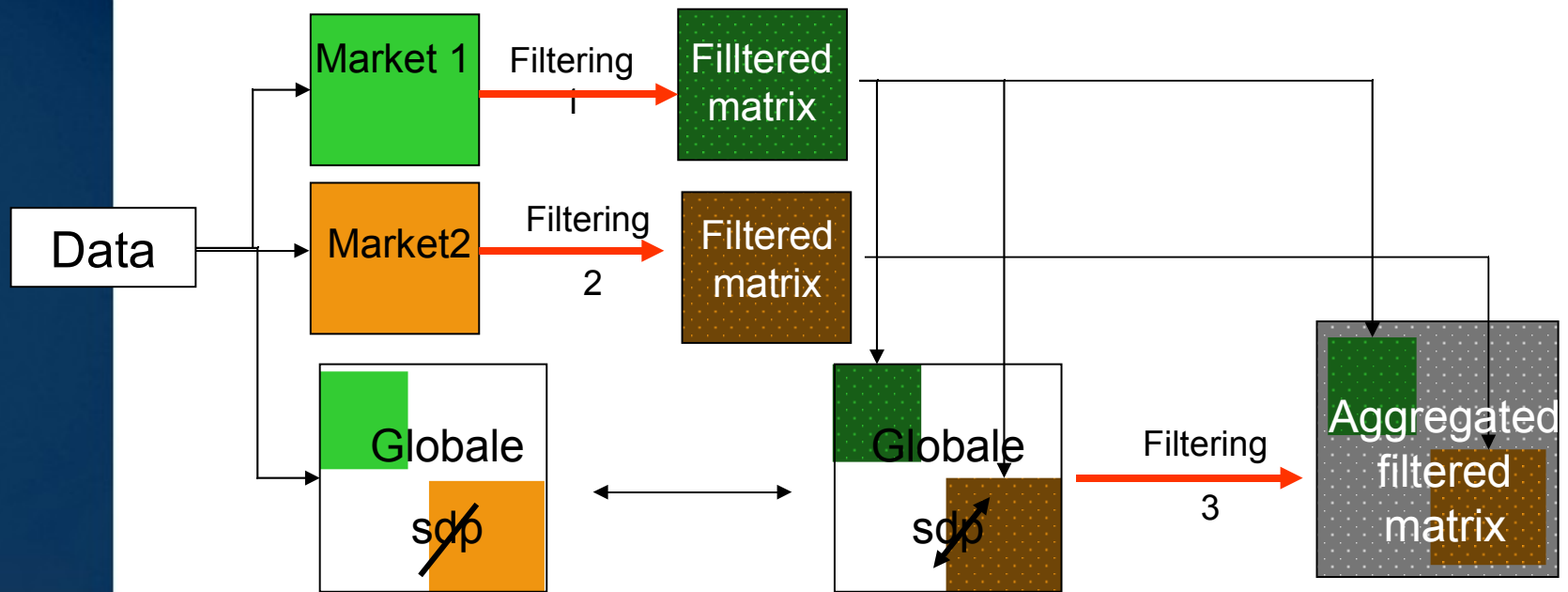
## Additional situations where filtering is needed

- Heterogeneous data series (asynchronicity due to large and diversified investment universes, multi data-provider...)
- Aggregation of local information into a global risk model (estimation of local matrices by independent statistical schemes)
- stress testing risk matrices (risk management)
- Prior views or intervals constraints on your matrix (volatilities, subset of the whole universe)

→ Destroying specific structural properties of risk matrices

 Pape Momar Ndiaye, François Oustry, Véronique Piolle. *Semidefinite optimisation for global risk modeling*, Journal of Asset Management 7, 142-153, 2006".

## Example 1: Managing different markets risk models



## Example 2: Robust Estimation of the Correlation Risk

### Context:

- Estimation of the correlation risk for derivative products over a basket of underlyings
- Multi-asset Monte Carlo simulations (via Cholesky factorization)

### Process:

- Stress testing the correlation matrix (applying a well chosen perturbation: direction, intensity)

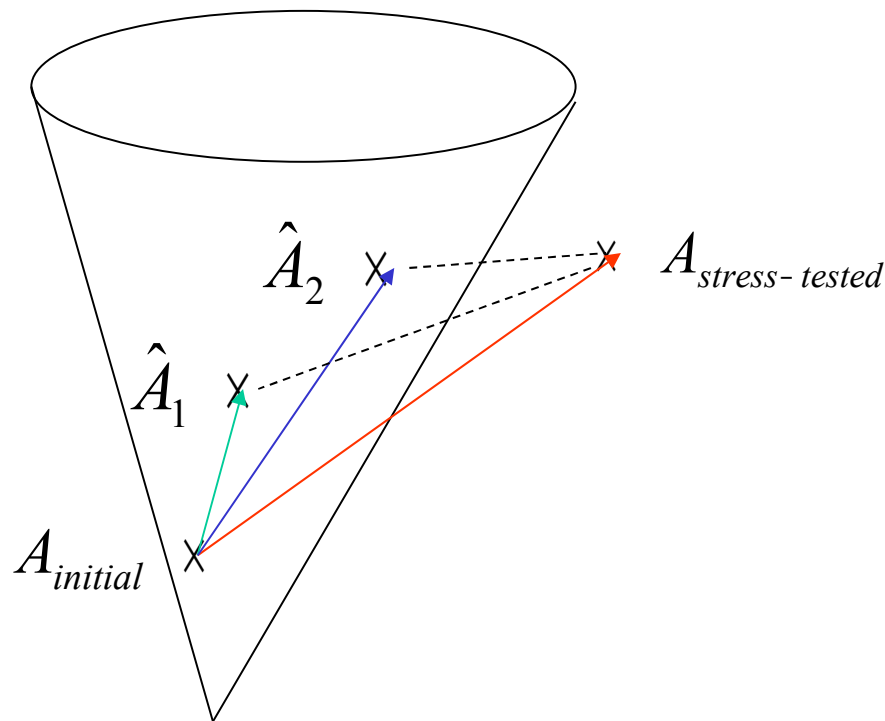
### Drawback:

- The perturbed matrix is not a correlation matrix anymore
- It cannot feed a simulation scheme or an optimization problem!!

### What to do?

- Cautiously restoring the PSD property
- Otherwise risk can be underestimated

## Example 2: Robust Estimation of the Correlation Risk



- Intensity and direction of the perturbation
- Intensity and direction of the perturbation taken into account with a rough correction
- Intensity and direction of the perturbation taken into account with a robust correction

## Why filtering matrices?

### Why underestimating small risk contributions is dangerous?

- Creating the illusion of portfolios with hedge ex ante sharpe ratios (factors linked to smallest eigenvalues)
- Then optimization will be driven by these noisy portfolios generating large turnovers and operational risks
- In fact, sensitivity analysis predicts this large turnovers

$$\min_{\omega} \omega^t \Gamma \omega$$

$$\begin{cases} \rho^t \omega \geq \ell \\ \sum_i \omega_i = 1 \end{cases}$$

(proportional to  $\frac{1}{\lambda_{\min}}$ )

A small perturbation of  $\rho$  or  $\Gamma$  is magnified at the portfolio/decision level by a factor proportional to  $\frac{1}{\lambda_{\min}}$ .

 L. El Ghaoui, F. Oustry and H. Lebret.

*Robust Solutions to Uncertain Semidefinite Programs SIAM J. Optimization, volume 9, no. 1, 1998.*



## Frequently used practices

### Increasing the smallest eigenvalues, restoring unit diagonal (correlation matrix)

- The sorted spectrum of the correlation matrix  $(\lambda_i)_{1 \leq i \leq n}$  and its corresponding eigenvectors matrix:  $(U)$
- Filtering level  $\sigma^2$

1/ Calibrating the correlation matrix	2/ Coming back to a correlation matrix
$D_{cor} = \begin{bmatrix} \lambda_1 & & & (0) \\ & \ddots & & \\ & & \lambda_{p-1} & \\ (0) & & \sigma^2 & \ddots \\ & & & & \sigma^2 \end{bmatrix}$ $X_{Cor} = U D_{Cor} U^T$	$V = \begin{bmatrix} \ddots & & & (0) \\ & & & \\ & & 1/\sqrt{X_{Cor ii}} & \\ (0) & & & \ddots \end{bmatrix}$ $X = V X_{Cor} V$

- ➔ Do not enable any other constraints
- ➔ Do not explicitly control conditioning number of final matrix neither the distance to the initial matrix

## Frequently used practices

### Factorial approach

“Does a factorial model improve (decrease) your condition number?”

Yes ...

Modeling the factorial approach:

- $r_i \approx \alpha_i + \sum_{j=1}^q \beta_{ij} f_j + \varepsilon_i$  for  $i \in [1, n]$  with  $\varepsilon_i \sim \mathcal{N}(0, \sigma_i^2)$  and  $\text{cov}(r_i, \varepsilon_j) = 0$
- with  $B = (\beta_{ij})_{\substack{1 \leq i \leq n \\ 1 \leq j \leq q}}$  and  $\Delta = \text{diag}(\sigma_i^2)$  then  $\Gamma \approx B \Gamma_f B^T + \Delta = \hat{\Gamma}$   
Systemic risk    specific risk (diagonal)
- Notation:  $\Lambda(A) = \text{spectrum}(A)$

Von Neumann:  $B \Gamma_f B^T \succeq 0 \implies \hat{\Gamma} - \Delta \succeq 0 \implies \Lambda(\hat{\Gamma}) \geq \Lambda(\Delta) \implies \lambda_{\min}(\hat{\Gamma}) \geq \min_i \sigma_i^2$

... but The higher the explanatory power, the higher the condition number

➔ Trade-off between condition number and explanatory power:

## Frequently used practices

### The shrinkage method (à la Ledoit & Wolf)

“The trade-off between bias and estimator error”

$\Gamma$  , the sample covariance matrix (unbiased, high estimations error due to high dimension & lacks of data)

$\bar{\Gamma}$  , the asymptotic covariance matrix

$\Gamma_k$  , a structured estimate of the covariance matrix (biased)

(  $\eta I$  , the single-index covariance matrix, a k-factors matrix...)

The shrunked matrix is:  $\hat{\Gamma} = \alpha \Gamma_k + (1 - \alpha) \Gamma$

Where  $\alpha$  is the solution of 
$$\begin{cases} \min_{\alpha} \left\| \bar{\Gamma} - (\alpha \Gamma_k + (1 - \alpha) \Gamma) \right\|^2 \\ st \ \alpha \in ]0,1[ \end{cases}$$

Numerical approximation of the Shrinkage estimator  $\alpha$

## Frequently used practices

### The shrinkage method (à la Ledoit & Wolf)

Improving the condition number?

Von Neumann  $\rightarrow$  As far as  $\Gamma$  and  $\Gamma_k$  are SDP,  $\lambda_{\min}(\hat{\Gamma}) \geq \lambda_{\min}(\Gamma_k)$

$\rightarrow$  As far as  $\Gamma_k$  is well conditioned,  $\hat{\Gamma}$  as well.

How do we perform the filtering?

## Taking a conservative robust-optimization point of view

Preserving “**conservative**” requirements while producing “**reactive**” matrices

- We want to capture current trends
  - We want this matrix to produce robust portfolio decisions/deviations
- Reducing the estimation windows, incorporating implied information
- Filtering the noise due to the lack of data

*Like if you were designing a robust control law for an fighter aircraft, you need to react quickly but conservatively using very short term information through a very cautious filtering scheme*

**We need to depart from these best estimations  
to obtain ex-ante robustness\* certificates**

(\* robust-optimization meaning)

## Taking a conservative robust-optimization point of view

Calibration thanks to Semidefinite programming

$$\left\{ \begin{array}{l} \min_{\Gamma \in \Omega} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \\ X \succeq \sigma^2 I \end{array} \right. \begin{array}{l} \longleftarrow \text{Froebenius norm: } \text{Tr}((X - \Gamma)^2) \\ \text{very suited to capture the geometry} \\ \text{of calibration problem} \\ \longleftarrow \text{Linear constraints} \\ \longleftarrow \text{Restoring SDP + reconditioning} \end{array}$$

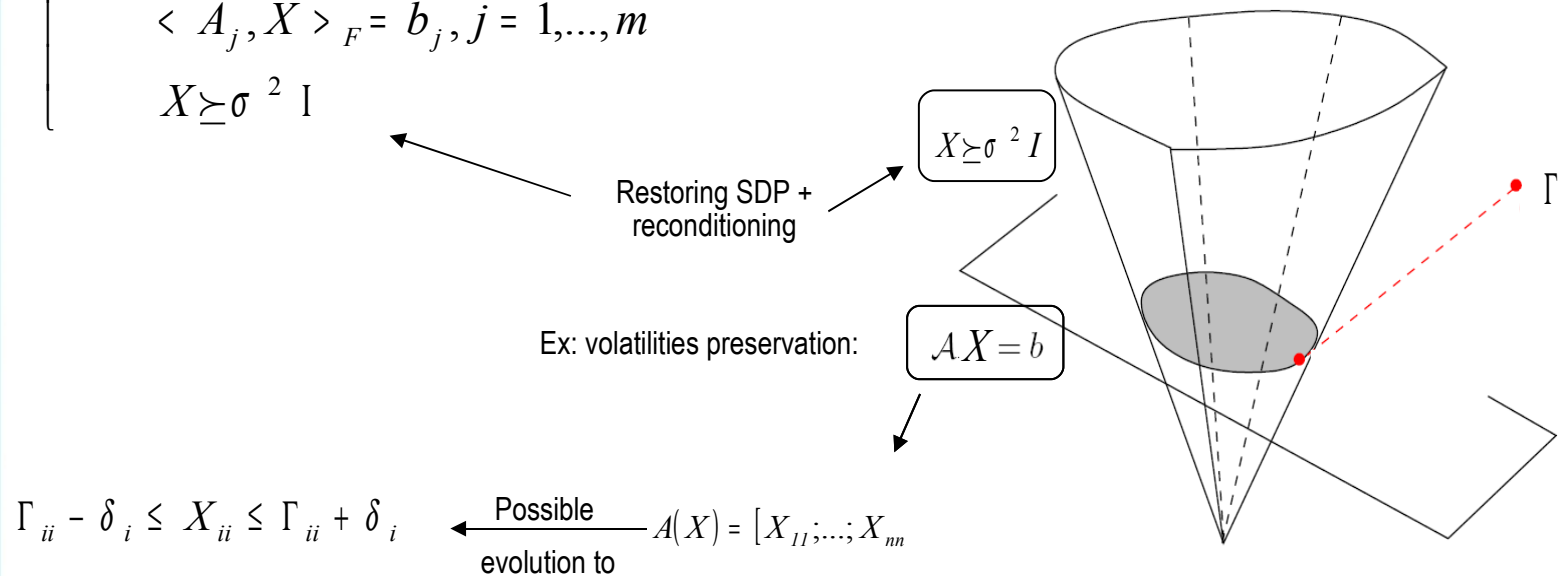
$$\begin{array}{c} \uparrow \\ X \succeq \sigma^2 I \Leftrightarrow X - \sigma^2 I \text{ SDP} \end{array} \quad \sigma^2 = \text{noise filtering level (lower bound for spectrum)}$$

There is no direction where the risk can be seen smaller than a level linked to  $\sigma^2$

## Taking a conservative robust-optimization point of view

A geometrical representation

$$\left\{ \begin{array}{l} \min_{\Gamma \in \Omega} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \\ X \succeq \sigma^2 I \end{array} \right.$$



# Taking a conservative robust-optimization point of view

A SDP formulation of the filtering problem:

$$\left\{ \begin{array}{l} \min_{\Gamma \in \Omega} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \\ X \succeq \sigma^2 I \end{array} \right.$$

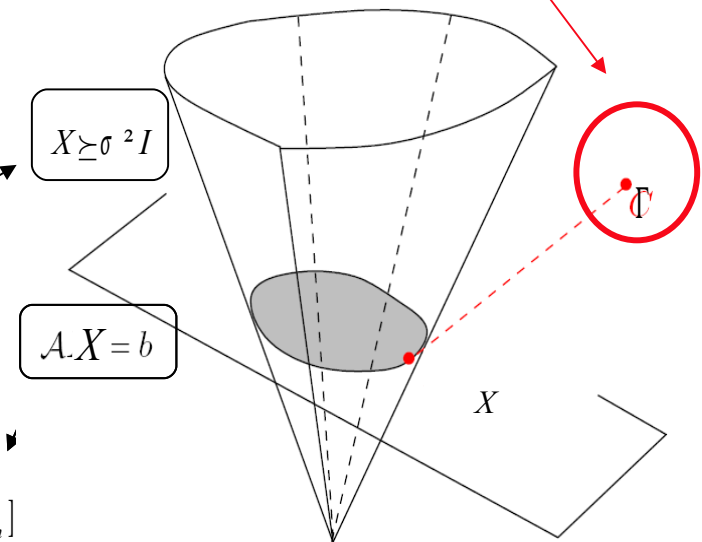
Level for the smallest eigenvalue

Filtering= Enforcing low condition number

Ex: volatilities preservation:

$$\Gamma_{ii} - \delta_i \leq X_{ii} \leq \Gamma_{ii} + \delta_i \xleftarrow{\text{Possible evolution to}} A(X) = [X_{11}; \dots; X_{nn}]$$

The original risk matrix, target



# Taking a conservative robust-optimization point of view

A SDP formulation of the filtering problem:

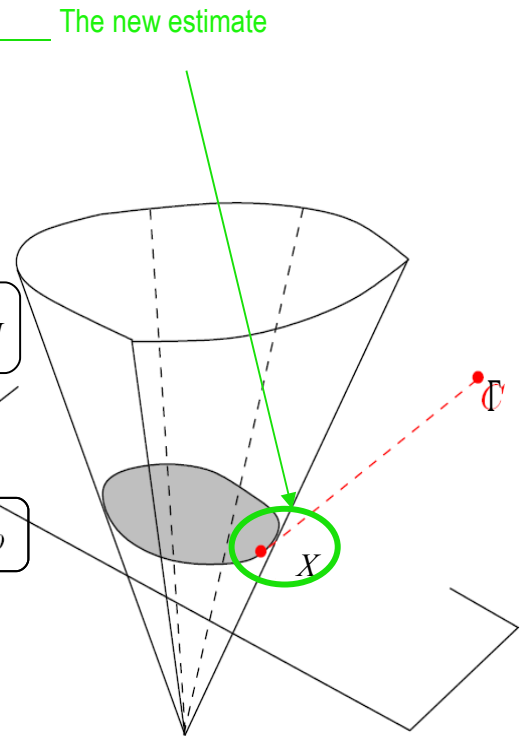
$$\begin{cases} \min_{X \in \Omega} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \\ X \succeq \sigma^2 I \end{cases}$$

Level for the smallest eigenvalue

Filtering= Enforcing low condition number

Ex: volatilities preservation:

$$\Gamma_{ii} - \delta_i \leq X_{ii} \leq \Gamma_{ii} + \delta_i \xleftarrow{\text{Possible evolution to}} A(X) = [X_{11}; \dots; X_{nn}]$$



## Taking a conservative robust-optimization point of view

A SDP formulation of the filtering problem:

$$\min_{\Gamma \in \Omega} \|X - \Gamma\|_F^2$$

subject to

$$\langle A_j, X \rangle_F = b_j, j = 1, \dots, m$$

$$X \succeq \sigma^2 I$$

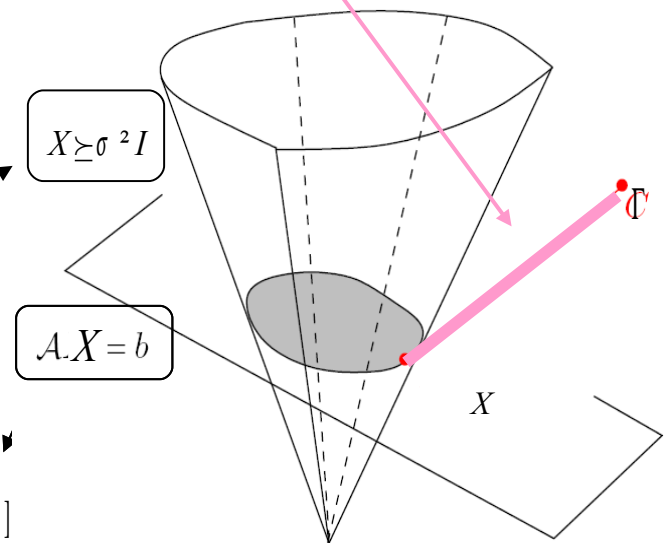
Level for the smallest eigenvalue

Filtering=  
Enforcing low condition number

Ex: volatilities preservation:

$$\Gamma_{ii} - \delta_i \leq X_{ii} \leq \Gamma_{ii} + \delta_i \xleftarrow{\text{Possible evolution to}} A(X) = [X_{11}; \dots; X_{nn}]$$

Minimizing the distance  
= projection



## Taking a conservative robust-optimization point of view

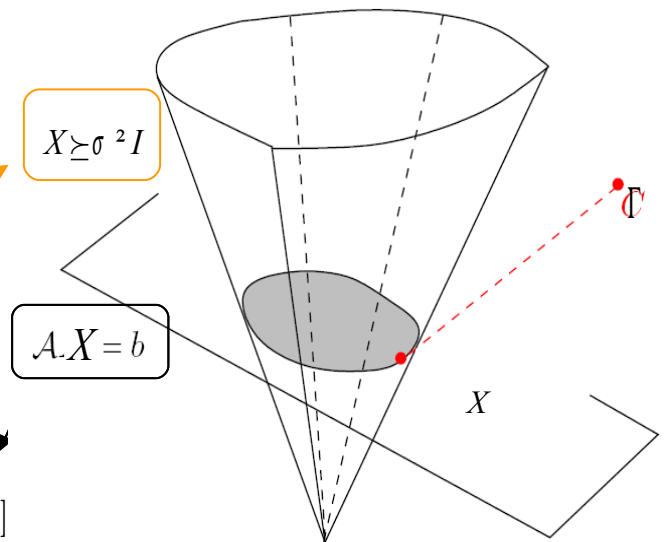
A SDP formulation of the filtering problem:

$$\left\{ \begin{array}{l} \min_{\Gamma \in \Omega} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \\ X \succeq \sigma^2 I \end{array} \right.$$

Level for the smallest eigenvalue

Filtering=  
 Enforcing low condition number

Ex: volatilities preservation:



$$\Gamma_{ii} - \delta_i \leq X_{ii} \leq \Gamma_{ii} + \delta_i \xleftarrow{\text{Possible evolution to}} A(X) = [X_{11}; \dots; X_{mm}]$$

## Taking a conservative robust-optimization point of view

A SDP formulation of the filtering problem:

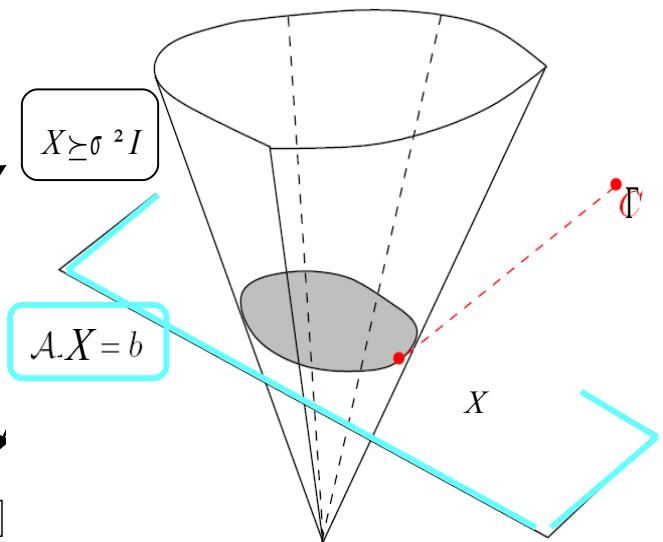
$$\left\{ \begin{array}{l} \min_{\Gamma \in \Omega} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \\ X \succeq \sigma^2 I \end{array} \right.$$

Level for the smallest eigenvalue

Filtering= Enforcing low condition number

Ex: volatilities preservation:

$$\Gamma_{ii} - \delta_i \leq X_{ii} \leq \Gamma_{ii} + \delta_i \xleftarrow{\text{Possible evolution to}} A(X) = [X_{11}; \dots; X_{mm}]$$



## Taking a conservative robust-optimization point of view

### Example of possible constraints

Preserving volatilities

Preserving the trace (whole universe risk level)

Preserving some of entries (better estimated than other)

Giving prior views or intervals on some of entries

Including implied information

## Algorithm

Primal pb

$$(primal): \begin{cases} \min \frac{1}{2} \|X - \Gamma\|_F^2 \\ \text{subject to} \\ X \in S_n^+ \\ \langle A_j, X \rangle_F = b_j, j = 1, \dots, m \end{cases}$$

Structure not easy to exploit

Dual pb more strongly structured

Lagrangian dualization (of linear constraints)

Linear operator:  $A(X) = \begin{bmatrix} \langle A_1, X \rangle \\ \vdots \\ \langle A_m, X \rangle \end{bmatrix}$

Partial Lagrangian:  $L(X, y) = \frac{1}{2} \|X - \Gamma\|_F^2 - \langle y, A(X) - b \rangle$

Dual function:  $\theta(y) = \min_{X \in S_n^+} L(X, y)$

## Algorithm

**Dual pb**  $(dual) \left\{ \max_{y \in \mathfrak{R}^m} \theta(y) \right\} \longleftrightarrow \left\{ \min_{y \in \mathfrak{R}^m} \frac{1}{2} \left\| P_{S_n^+}(\Gamma + A^* y) \right\|^2 - b^T y \right\}$

## Properties of dual function

Concave

Differentiable

Gradient Lipschitz continuous

## Why?

- The dual function reaches its min for a single value i.e. :

$$\partial \theta(\lambda) = \underset{x \in S_n^+}{\text{Argmin}} L(x, \lambda) \text{ is a singleton and } \theta \text{ is differentiable}$$

- Gradient Lipschitz continuous because  $\theta$  can be seen as a Moreau-Yoshida regularization function

## Algorithm

**Dual pb**      (dual)  $\left\{ \max_{y \in \mathfrak{R}^m} \theta(y) \right\} \longleftrightarrow \left\{ \min_{y \in \mathfrak{R}^m} \frac{1}{2} \left\| P_{S_n^+} (\Gamma + A^* y) \right\|^2 - b^T y \right\}$

## Resolution of the dual pb

Quasi-Newton algorithm type  $\rightarrow y^*$

**Resolution of primal pb**  $X^* = p_{S_n^+} (\Gamma + A^* y^*)$

$p_{S_n^+}$  is the projection onto the set  $S_n^+$

 Malick J. (2004), *A Dual approach to Semidefinite Least-Squares Problems*, SIAM Journal of Matrix Analysis and Applications Vol. 26(1) pp. 272-284.

## Example of filtering impact on an investment strategy

- **Context**

Dynamic overlay to be used to boost a monetary fund

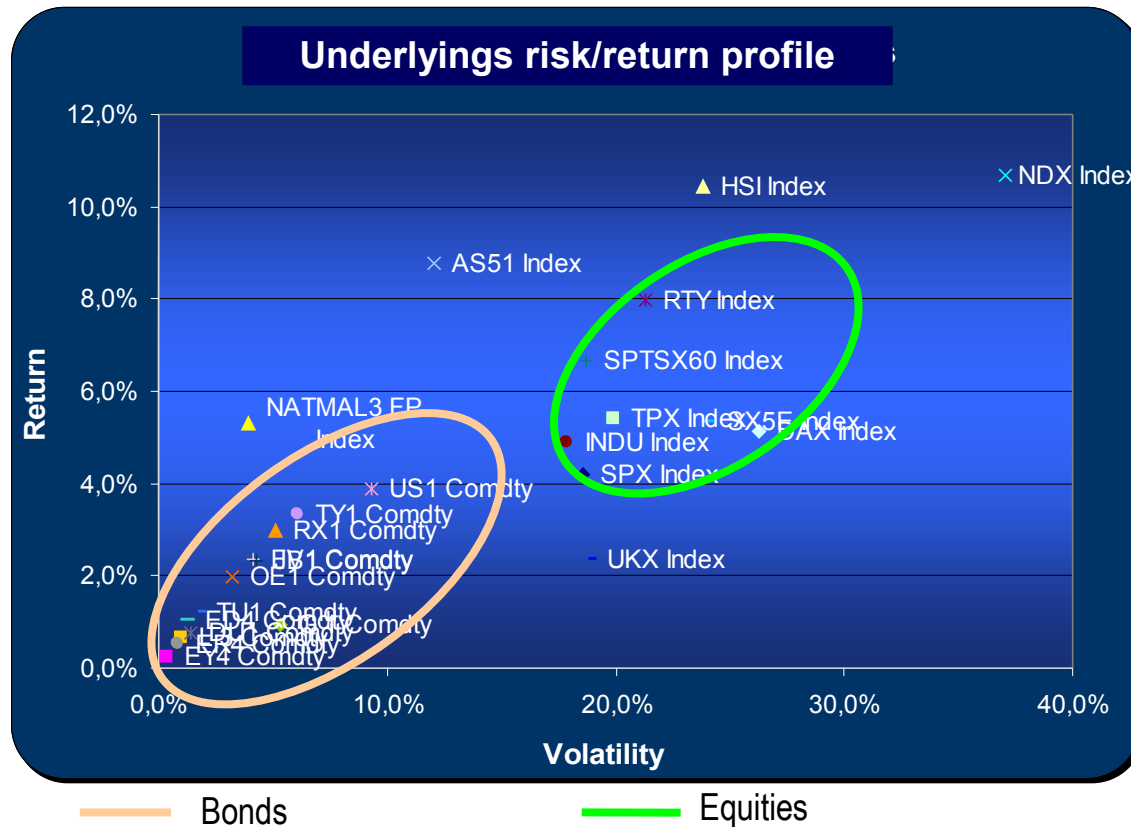
Back-test from October 1998 to October 2006

Allocation frequency: 3 months

- **Investment universe**

- Futures on equities and bonds + one cash position
- Range of annual volatilities:
  - 1 to 10% for bonds
  - 10 to 40% for equities

## Example of filtering impact on an investment strategy



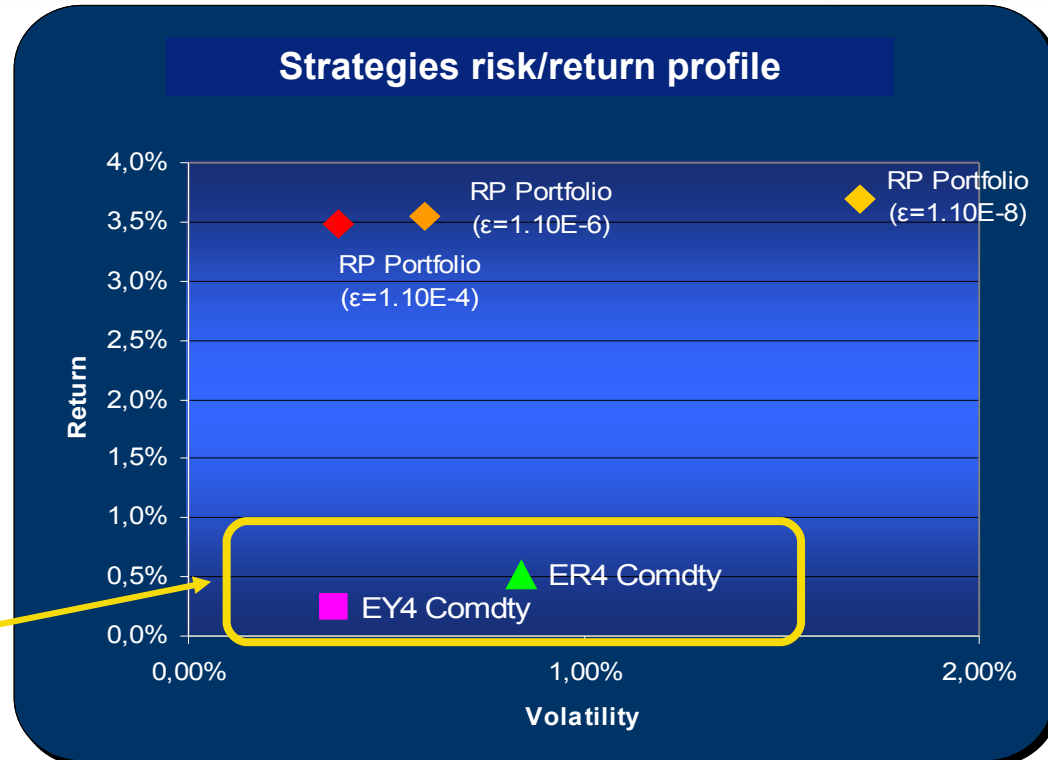
## Example of filtering impact on an investment strategy

### Presentation of the investment strategy

- Allocation model:  
Maximization of return under  
volatility budget constraint
- Return model : internal scoring model
- Risk model: 3 filtering level were tested ( $10^{-8}$ ,  $10^{-6}$ ,  $10^{-4}$  ~ daily variances)  
( $10^{-8}$  ~ 1bp on daily vol or 15 bp on annualized vol)
- Criteria to determine the performance of the strategy:
  - Turn over and max rate of trading
  - Saturation of ex post volatility budget
  - Standard ratio

$$\left\{ \begin{array}{l}
 \text{Max } \rho^T \omega + \rho_0 \omega_0 \\
 \text{s.t. } Vol_{Ex-Ante}(\omega, \omega_0) \leq 0,5\% \\
 -1,1 \leq \omega_i \leq 1,1 ; \text{ For futures on equities} \\
 -2 \leq \omega_i \leq 2 ; \text{ For futures on bonds} \\
 -1 \leq \omega_0 \leq 1 ; \text{ For cash} \\
 \sum_{i=0}^n \omega_i = 1
 \end{array} \right.$$

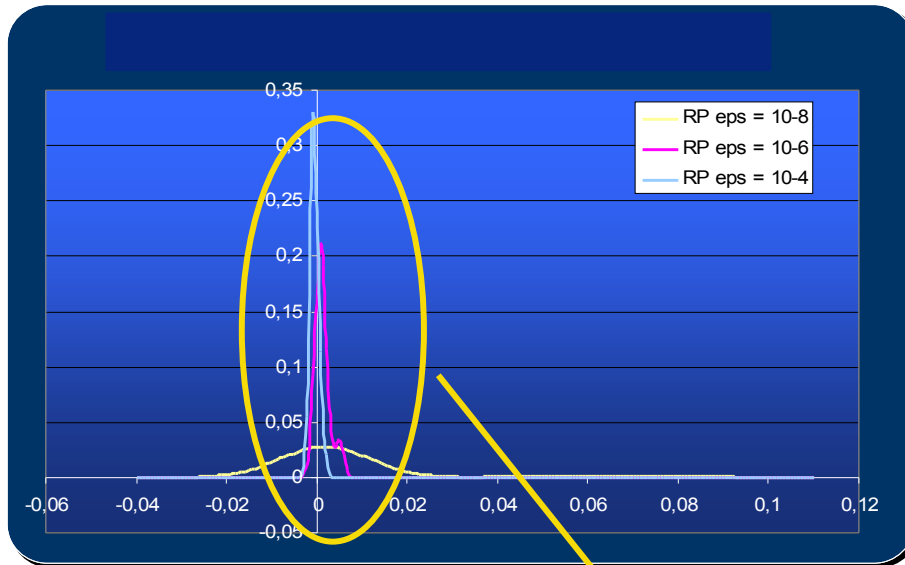
## Example of filtering impact on an investment strategy



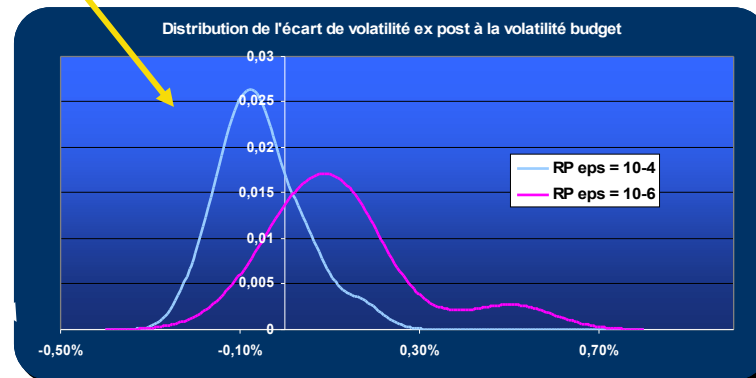
Underlyings in the same volatility range as strategies

	RP eps = 10-8	RP eps = 10-6	RP eps = 10-4
<b>Return</b>	<b>3,70%</b>	<b>3,54%</b>	<b>3,52%</b>
<b>Volatility</b>	<b>1,70%</b>	<b>0,60%</b>	<b>0,46%</b>
<b>Gross ratio</b>	<b>2,18</b>	<b>5,93</b>	<b>7,72</b>

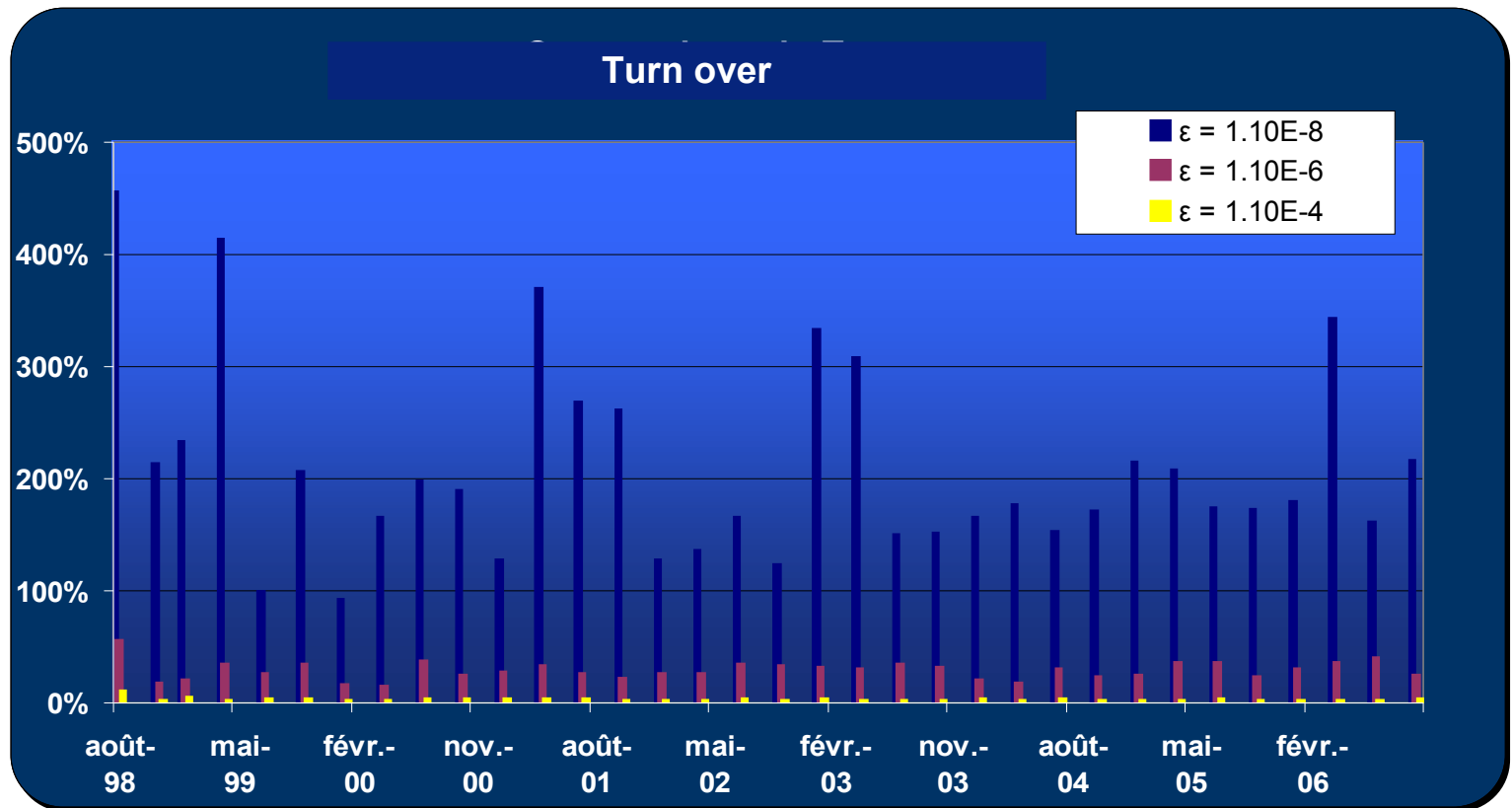
## Example of filtering impact on an investment strategy



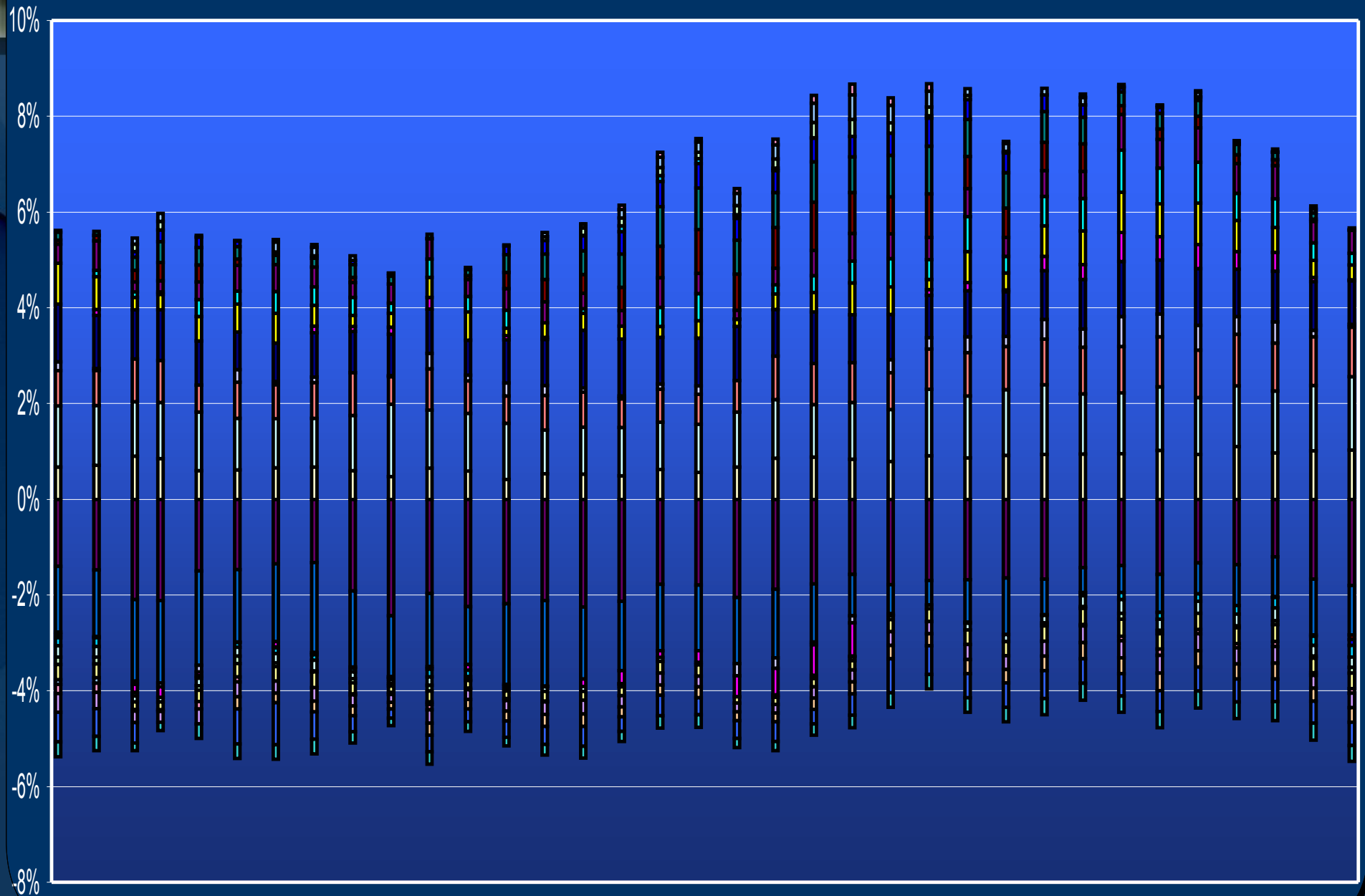
Distribution of deviation from ex post volatility to volatility budget  
 → Filtering reduces the gap ex ante ex post of volatility budget and is more conservative



## Example of filtering impact on an investment strategy



# Rolling pie of long/short positions without cash



## Going further: real life investment strategy

