# Performance assessment of optimal allocation for large portfolios

Luigi Grossi and Fabrizio Laurini

`luigi.grossi@univr.it` `fabrizio.laurini@unipr.it`

Dipartimento di Economie, Societá e Istituzioni, Universitá di Verona

Dipartimento di Economia, Universitá di Parma

# Basic idea of this talk

- Idea: Having got 10,000 euros, we want to find an optimal way of investing this money into a set (say $N$) of given shares in an optimal way as to maximize our expected returns and minimize risks after a fixed investment period.

- This investment should be:

    - robust to estimation errors,
    - stable over time.

# Definitions

Let $\boldsymbol{P}_t = (P_{1t}, \dots, P_{Nt})$ be the closing price of $N$ stocks on time $t$.

Let $\boldsymbol{y}_t = log(\boldsymbol{P}_t/\boldsymbol{P}_{t-1})$ be 1-period returns of stocks. We assume that the period return $\boldsymbol{y}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

Let $\boldsymbol{x} = (x_1, ..., x_N)$ denote the share of our wealth in stocks (vector of weights), such that $\sum_{i=1}^{N} x_i = 1$.

Let $\mu_p = \boldsymbol{x}'\boldsymbol{\mu}$ and $\sigma_p^2 = \boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x}$ the expected return and variance of the portfolio, respectively.

# Mean-variance optimization procedure

In traditional mean variance optimization of a portfolio with $N$ assets, expected utility is maximized using the following Lagrangian function

$$L = \boldsymbol{\mu}'\boldsymbol{x} - \frac{1}{2}\gamma\boldsymbol{x}'\boldsymbol{\Sigma}\boldsymbol{x} - \lambda(\boldsymbol{x}'\boldsymbol{1} - 1)$$

where
$\gamma$: is the parameter of relative risk aversion,
$\lambda$: is a Lagrange multiplier.

R1: $(\boldsymbol{x}'\boldsymbol{1} - 1)$ is the constraint requiring that the optimal weight vector's elements sum up to 1. Sometimes no short selling constraint is imposed, that is $x_i \geq 0, \forall i$.

R2: $\boldsymbol{\mu}$ and $\Sigma$ are estimated by ML.

R3: Sensitivity to "outliers".

# The problem

- Traditional Mean-Variance optimization (Markowitz) assumes known expected returns $\mu$ and covariance matrix $\Sigma$

- In practice, $\mu$ and $\Sigma$ must be estimated and therefore contain estimation error.

- Outline
  - simulation to show the impact of estimation error on optimal asset allocation.
  - new robust estimators of portfolio mean-variance and comparison with other robust estimators (basically MCD).
  - application to real data.

# Simulation experiment

$N = 6$, with true given parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. $T = 120$, 10-years time series.

Uncontaminated series

$$\boldsymbol{y}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Contaminated series

$$\boldsymbol{y}_t^* = \boldsymbol{y}_t + \theta \boldsymbol{\delta}_t$$

where $\theta > 0$ is the magnitude of the Additive Outlier and $\boldsymbol{\delta}_t$ is a stochastic contamination process.

# True, actual and estimated frontiers

Efficient frontier is obtained linking points on the mean-variance plot. Different types of frontiers can be drawn to compare estimators.

- True efficient frontier

$$\mu_p = x'\mu \quad \sigma_p^2 = x'\Sigma x$$

- Estimated frontier

$$\mu_p = \hat{x}'\hat{\mu} \quad \sigma_p^2 = \hat{x}'\hat{\Sigma}\hat{x}$$

- Actual frontier

$$\mu_p = \hat{x}'\mu \quad \sigma_p^2 = \hat{x}'\Sigma\hat{x}$$

# True, actual and estimated frontiers



True, actual and estimated frontiers

# **Uncontaminated data**

True and actual frontiers with MLE.

**True frontier (solid black) and non−contaminated actual frontiers**

# Contaminated data

True and actual frontiers with MLE.

**True frontier (solid black) and contaminated actual frontiers**
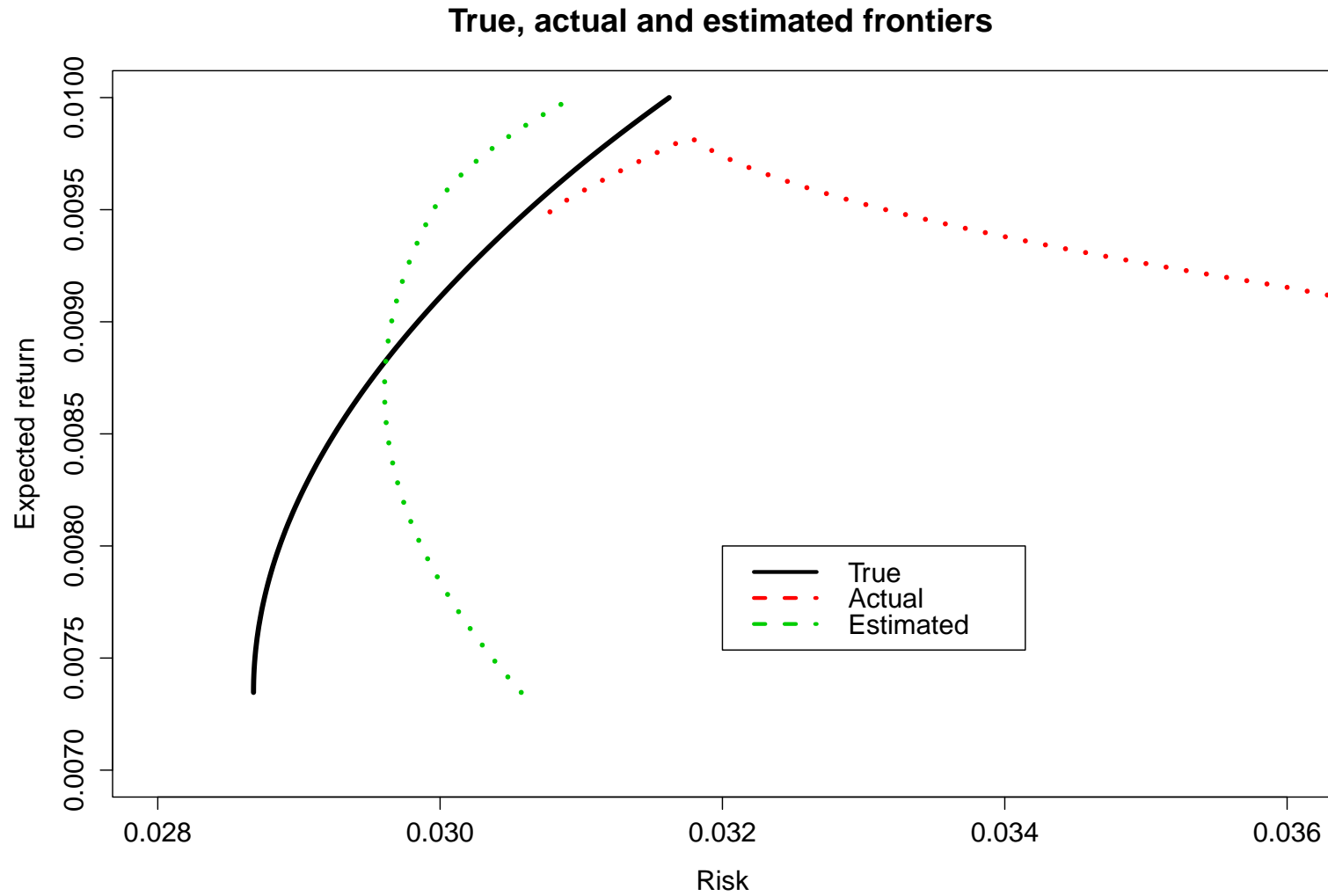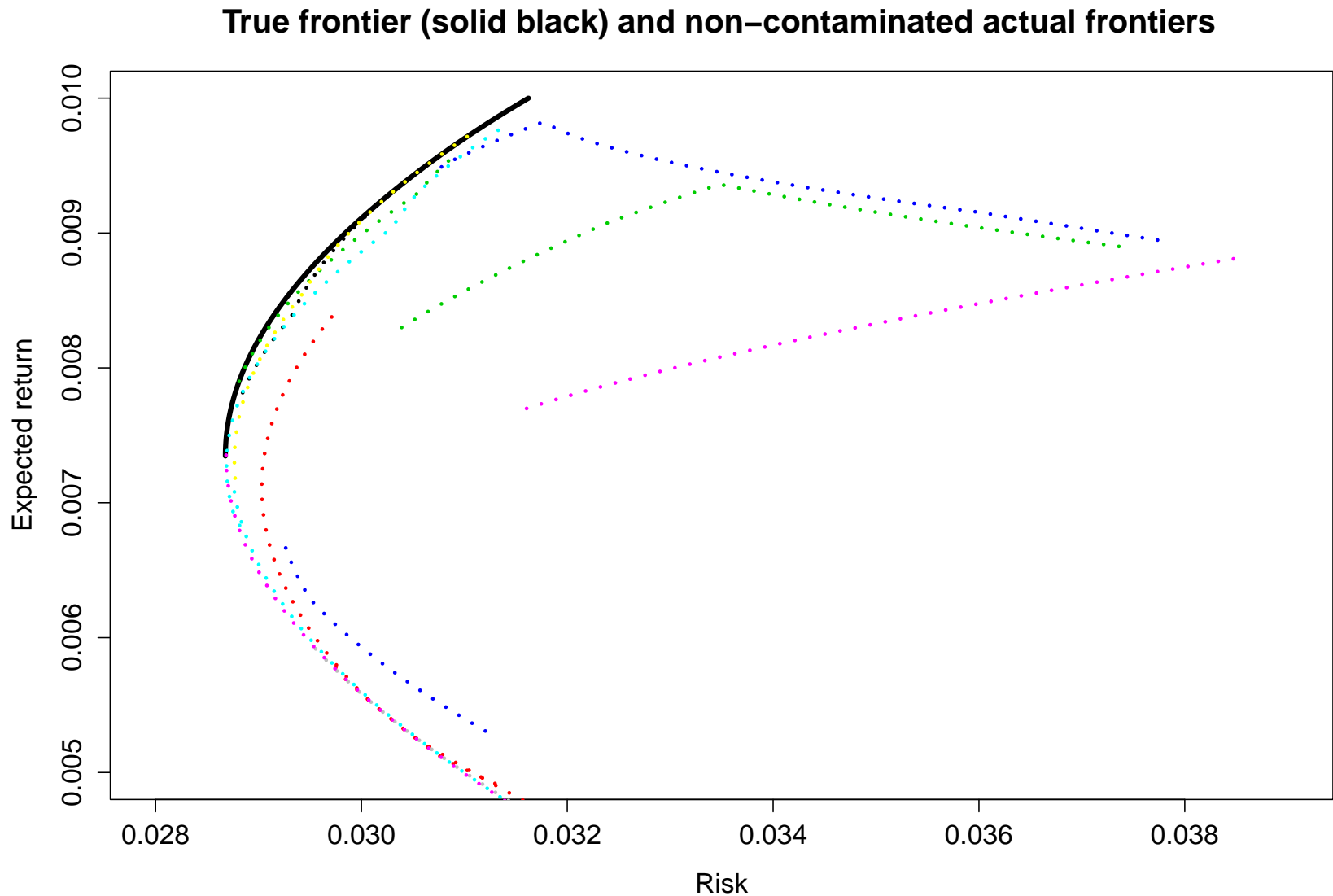
# Assessing estimation errors

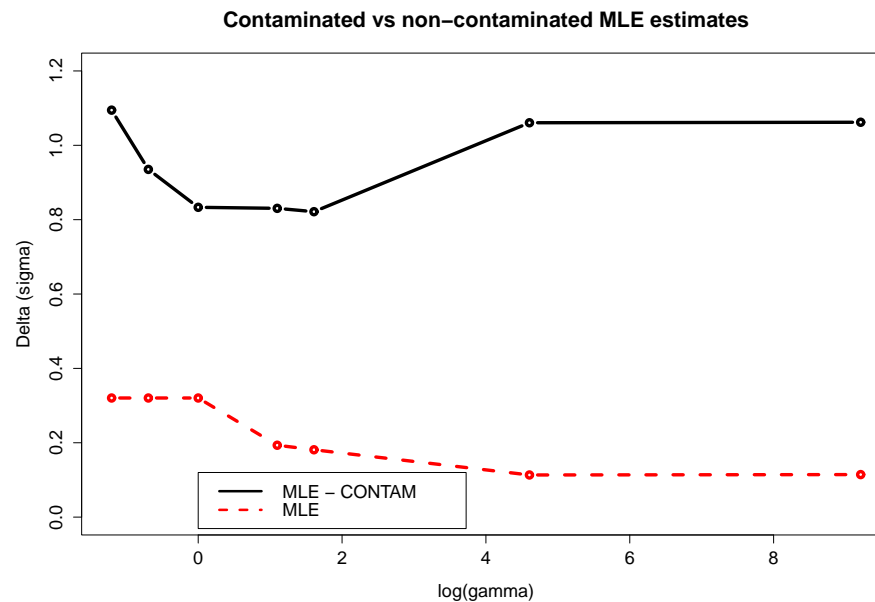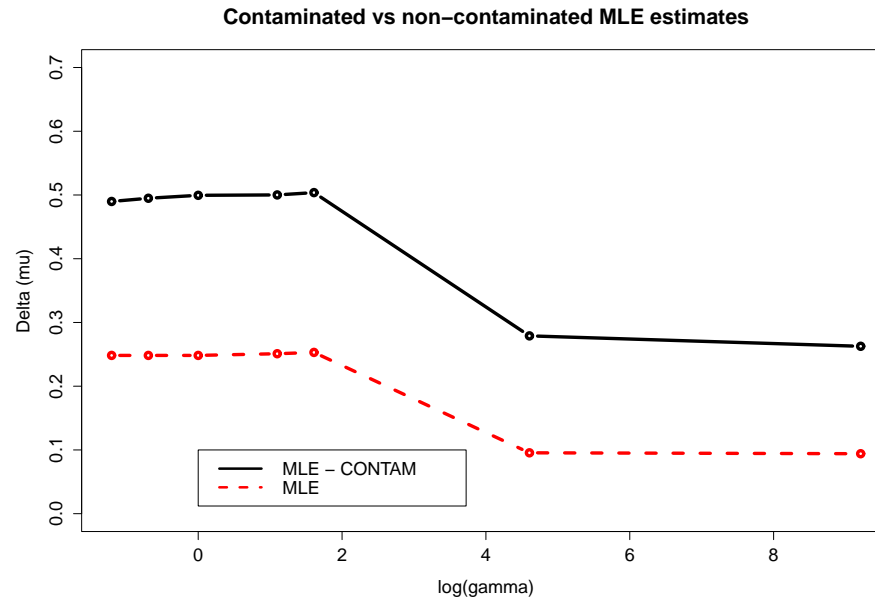Quantitative measure of the estimation error using a given estimator.

$$\Delta_\mu(\gamma) = \sqrt{\frac{1}{S}\sum_{i=1}^{S}[\mu_p(\gamma) - \tilde{\mu}_p^s(\gamma)]^2} \quad \Rightarrow RMSE \text{ for } \mu_p$$

$$\Delta_\sigma(\gamma) = \sqrt{\frac{1}{S}\sum_{i=1}^{S}[\sigma_p(\gamma) - \tilde{\sigma}_p^s(\gamma)]^2} \quad \Rightarrow RMSE \text{ for } \sigma_p$$

where $s = 1, \ldots, S$; $S$=number of simulations; $(\mu_p^s(\gamma), \sigma_p^s(\gamma))$ is the target point on the true frontier (given $\lambda$) maximizing the function $\mu_p - \gamma\sigma_p$.

# RMSE for MLE

RMSE for $\mu$ and for $\sigma$ as $\gamma$ increases. 1000 simulations.



Contaminated vs non−contaminated MLE estimates



Contaminated vs non−contaminated MLE estimates

# A new robust estimator

Target: to get a new robust version of the covariance matrix where observations are weighted according to their degree of outlyingness.

Method: distance of the Mahalanobis trajectories during the forward search and a distribution percentile

$$d_t^2 = (\boldsymbol{y}_t - \hat{\boldsymbol{\mu}})' \hat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{y}_t - \hat{\boldsymbol{\mu}})$$

where $\hat{\boldsymbol{\mu}}$ and $\hat{\Sigma}$ are MLE obtained using $T$ observations $\Rightarrow$ sensitivity to extreme observations. Multiple outliers could be masked.

# Forward search Mahalanobis distances

$$d_{(t)}^{*2} = (\boldsymbol{y}_t - \hat{\boldsymbol{\mu}}_m^*)'(\hat{\boldsymbol{\Sigma}}_m^*)^{-1}(\boldsymbol{y}_t - \hat{\boldsymbol{\mu}}_m^*)$$

where $\hat{\boldsymbol{\mu}}_m^*$ and $\hat{\boldsymbol{\Sigma}}_m^*$ are the mean and covariance matrix estimated on a $m$-sized subset with $m < T$

**R1:** observations belonging to the subset $S_*^{(m)}$ are selected according to the forward search criteria.

**R2:** extreme multivariate observations are added to the subset in the very last steps of the procedure.

**R3:** inclusion of extreme observations are pointed out by sudden increase/descrease of forward search trajectories. Influential observations show large values of $d_{(t)}^{*2}$ during the forward search.

# Assessing the degree of outlyingness

Distribution of Mahalanobis distance when $m = T$ (Riani, Atkinson and Cerioli, 2009):

$$d_{(t)}^{*2} \sim [T/(T-1)][N(m-1)/(m-N)]F_{N,T-N}$$

Target: to get weights $w_t \in [0,1]$ for $t = 1, \ldots, T$ to compute a weighted covariance matrix, such that the most outlying observations are down-weighted. In practice, a series of weighted returns are obtained, that is:

$$y_{it}^{\star} = y_{it}w_t^{1/2}.$$

The weighted covariance matrix will be simply the covariance matrix based on the weighted observations, which is denoted as $\tilde{W}$.

# Weighted covariance matrix

The procedure:

- measuring the distance between $d_{(t)}^{*2}$ and the quantile $\pm F_\delta$, as follows:

$$\pi_m^{(t)} = \begin{cases} 0 & \text{if } d_{(t)}^{*2} \in [0, F_\delta], \\ (d_{(t)}^{*2} - F_\delta)^2 & \text{if } d_{(t)}^{*2} > F_\delta, \end{cases}$$

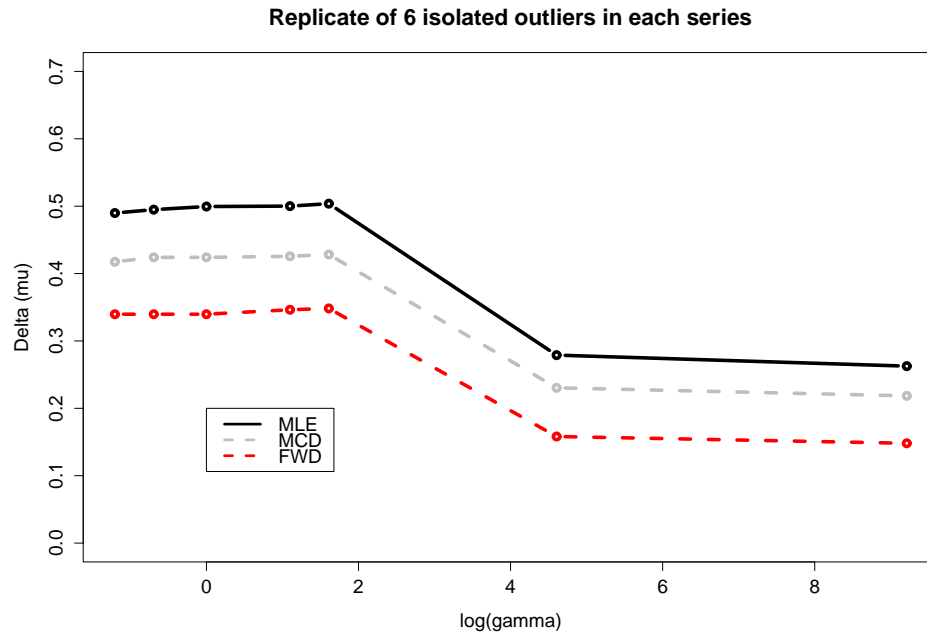- getting the overall distance of the $t$-th obervation

$$\pi_t = \frac{\sum_{m=m_0}^{T} \pi_m^{(t)}}{T - m_0}.$$

- mapping the distance

$$w_t = \exp(-\pi_t) \in [0, 1].$$

# Back to the example on simulated data

RMSE ($\mu$ and $\sigma$) for MCD, MLE and FWD with contaminated data. Single outliers.

**Replicate of 6 isolated outliers in each series**



**Replicate of 6 isolated outliers in each series**
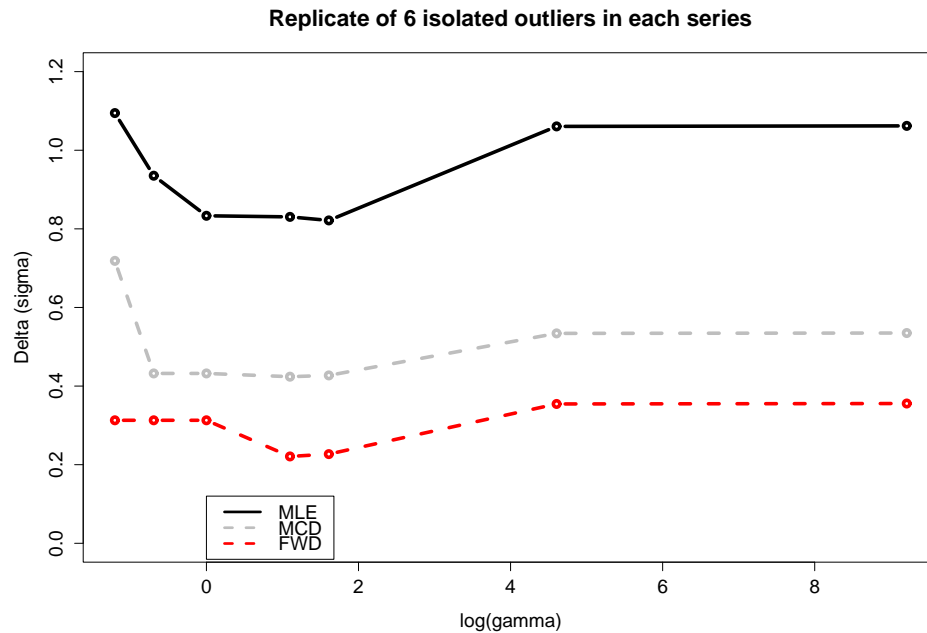
# Back to the example on simulated data

RMSE ($\mu$ and $\sigma$) for MCD, MLE and FWD with contaminated data. Patches of outliers.

**Patches of outliers in each series**



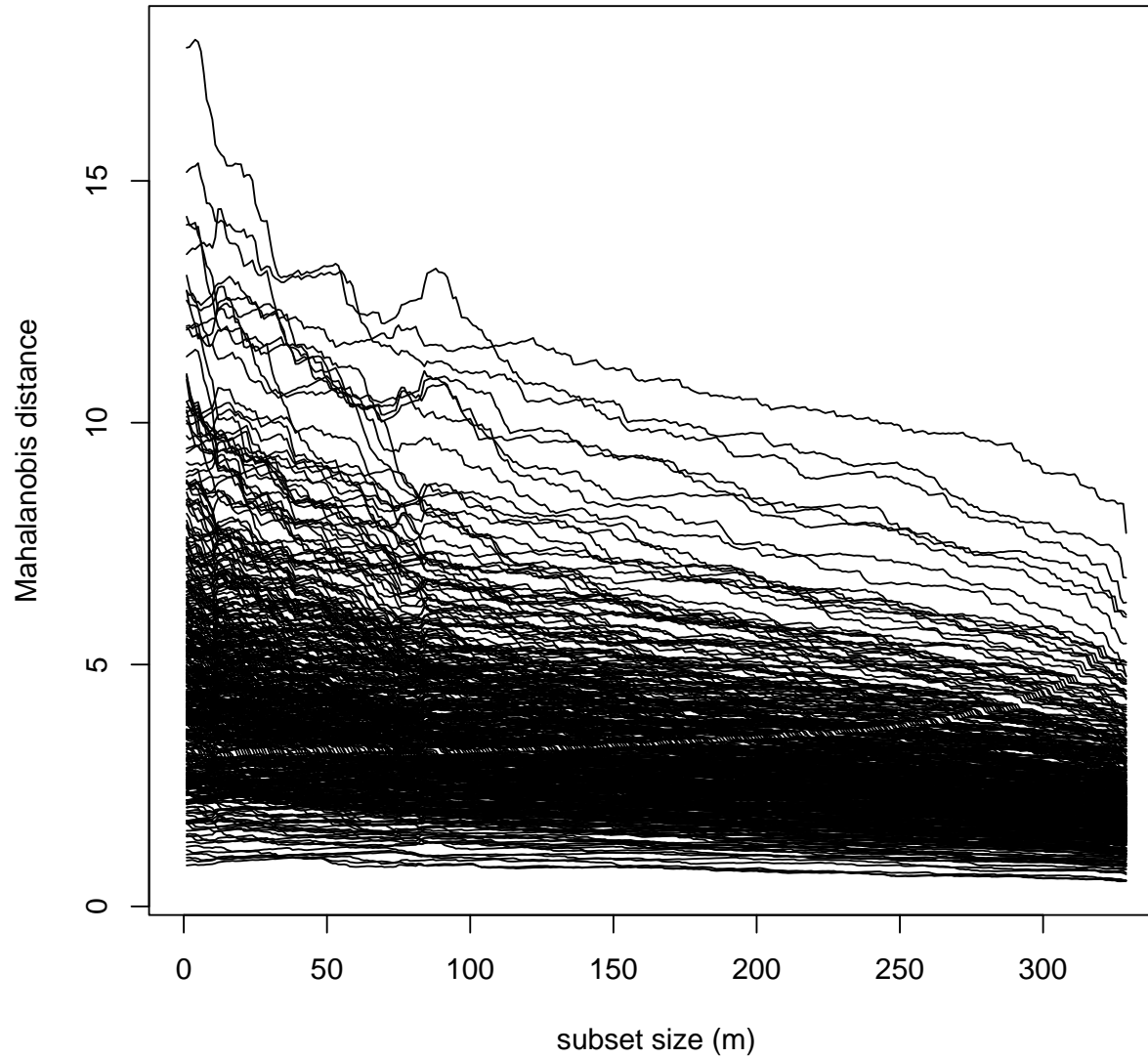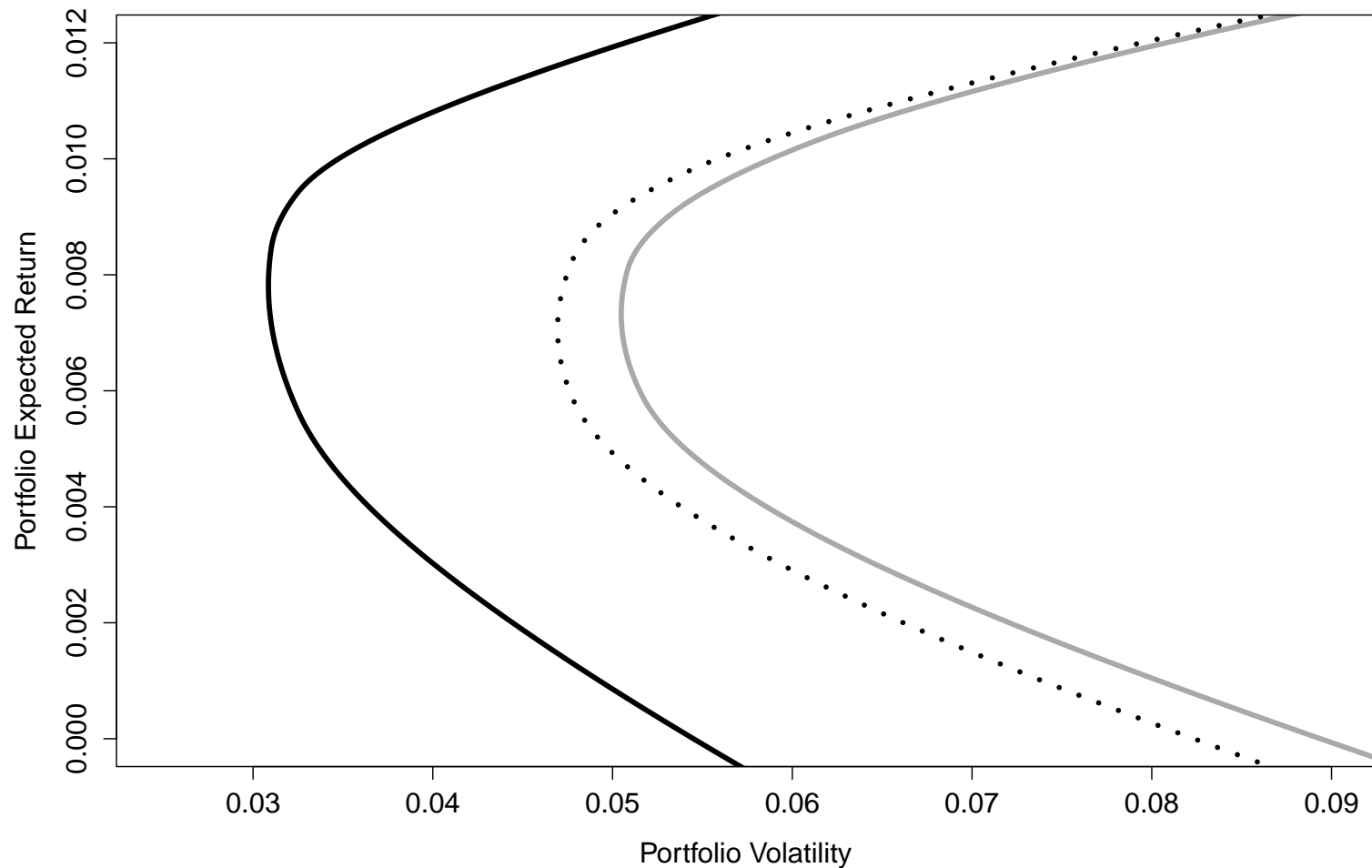**Patches of outliers in each series**
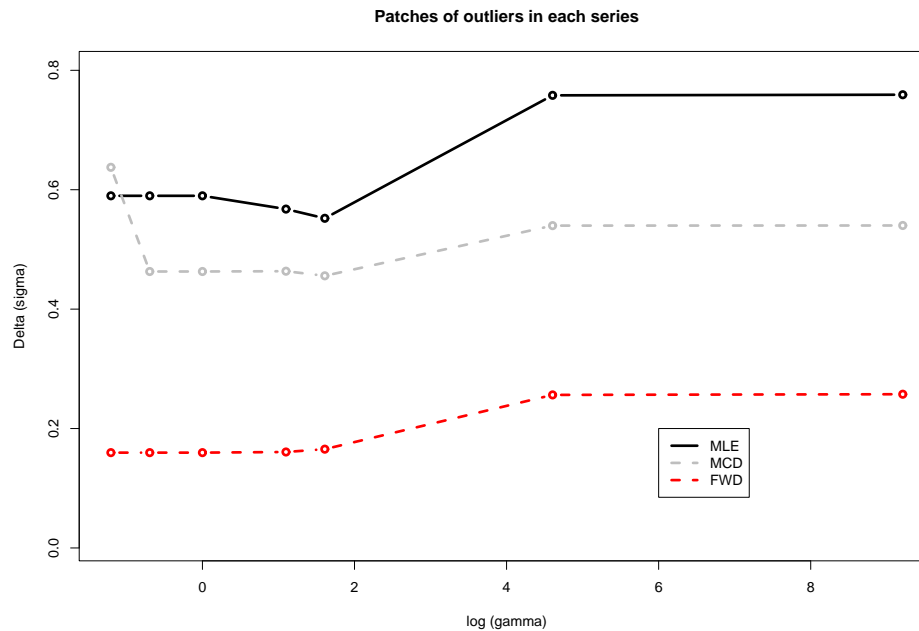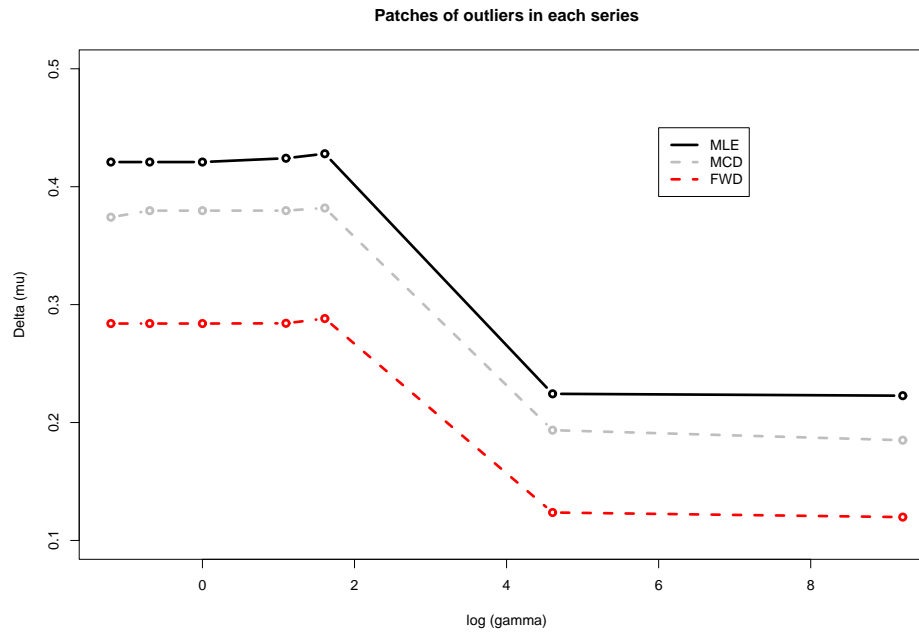
# US market stocks

Monthly returns of six stocks (Boeing, General Electric, General Motors, IBM, Procter and Gamble, Walt Disney) of the US market with data from January 1973 to March 2009 included. Data come from Datastream.
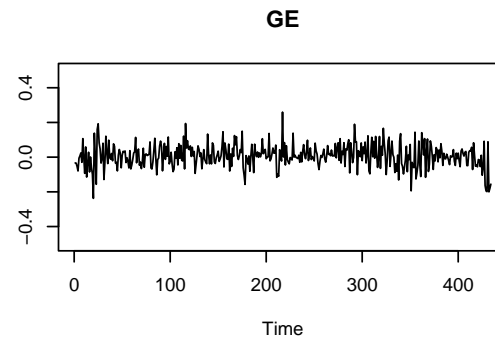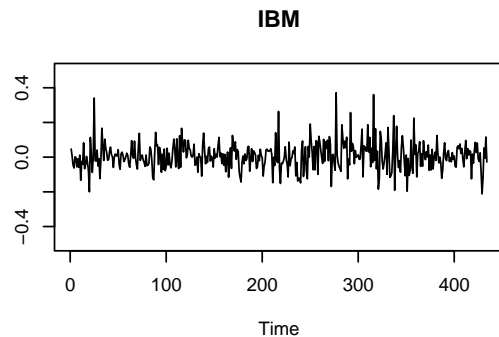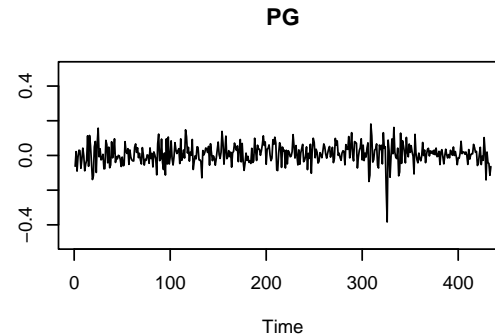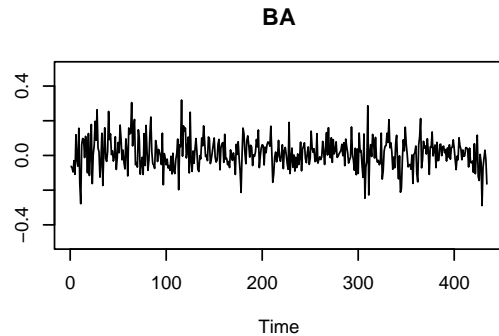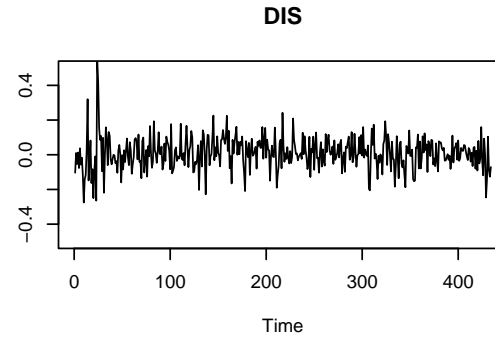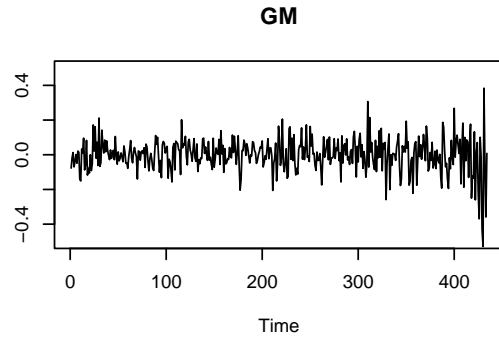
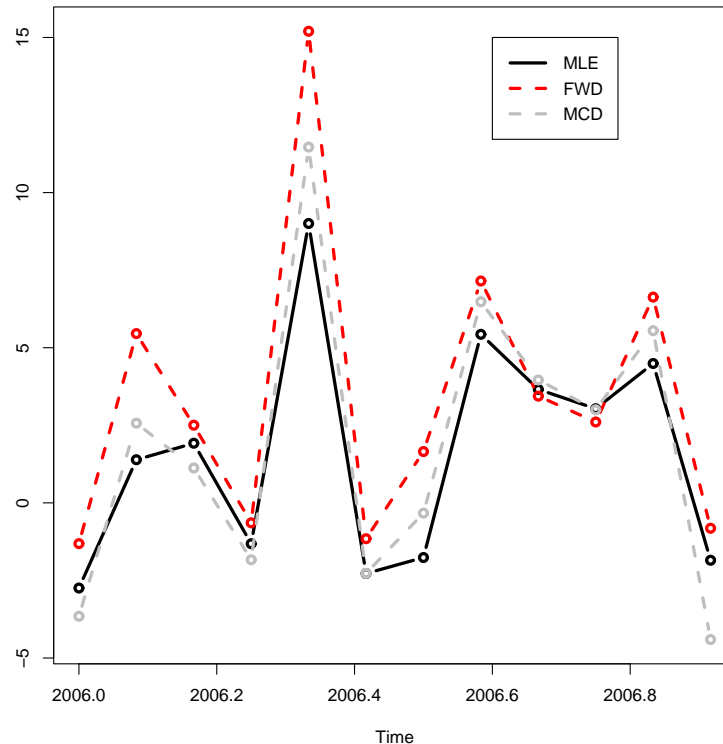# Mahalanobis distances trajectories

# Example - Efficient frontiers



Gray line: MLE Markowitz frontier. Black dashed line: estimated frontier using MCD robust estimator of the covariance matrix. Black solid line: estimated frontier through the forward weighted covariance matrix.

# Example - Portfolio returns



Portfolio performance using rolling windows in 2006. That is:

- estimate weights (MLE, MCD and FWD) using data for $t = 1, ..., T - 1$ and get the average portfolio return in $T$.

- estimate weights (MLE, MCD and FWD) using data for $t = 2, ..., T$ and get the average portfolio return in $T + 1$.

- ...

- estimate weights (MLE, MCD and FWD) using data for $t = d, ..., T - d$ and get the average portfolio return in $T - d + 1$.

# Final remarks and open issues

- robust estimator based on forward search performs well with simulated data;

- simulations with more general distributional assumptions;

- theoretical problems should be fixed (distribution of Mahalanobis distances during the search);

- new indexes to assess the superiority of robust methods in real applications.

# References

[1]  Atkinson, A. C., Riani, M. and Cerioli, A. (2004),
     *Exploring Multivariate Data with the Forward Search*.
     Springer–Verlag, New York.

[2]  Atkinson, A. C., Riani, M. and Cerioli, A. (2009),
     Finding an unknown number of multivariate outliers.
     *J.R. Statistical Society, Series B*, **71**(2), 447–466.

[3]  DeMiguel, V. and Nogales, F.J. (2009), Portfolio
     selection with robust estimation, *Operations
     Research*, doi: 10.1287/opre.1080.0566.

[4]  Fabozzi, F.J., Kolm, P.N., Pachamanova, D.A. and
     Focardi, S.M. (2007), *Robust Portfolio Optimization
     Management*, Wiley, New York.

[5]  Welsch, R.E. and Zhou, X. (2007), Application of
     robust statistics to asset allocation models. *Revstat.
     Statistical Journal*, **5**(1), 97–114.