Outlier Removal for Prediction of Covariance Matrices
With an Application to Portfolio Optimization

The Third International School on
Actuarial and Financial Mathematics
Feodisia Ukraine
5th of September 2000

Thomas Hellström
University of Umeå
Sweden

http://www.cs.umu.se/~thomash/
email: thomash@cs.umu.se

Overview of the talk

- Portfolio theory
- Predictions of Covariance matrices and return vectors
- The Naive Prediction
- An algorithm to remove outliers

Modern Portfolio Theory

Harry Markowitz, William Sharpe

- Investors hold portfolios of assets and are therefore focused on the return and risk for the whole portfolio, not for individual assets.
- The risk is quantified by the standard deviation for the portfolio.
- For a given expected return we can get different expected standard deviation depending on the mix of assets (due to varying correlations between the assets).
- The optimal portfolio can be determined by solving a quadratic optimisation problem.

Modern Portfolio Theory

- The stock prices are r.v. with returns $r_1, \ldots, r_n$ with :
  Expected values $\mu_1, \ldots, \mu_n \equiv R$
  Covariances $\sigma_{11}, \sigma_{12} \ldots, \sigma_{nn} \equiv C$
- The portfolio is defined by the weights $\mathbf{w} = w_1, \ldots, w_n$ and has:

a return $R_p = \sum_{i=1}^{n} w_i r_i$

with expectation $\mu_p \equiv ER_p = \sum_{i=1}^{n} w_i Er_i = \sum_{i=1}^{n} w_i \mu_i = w^T R$

and a variance $\sigma_p^2 = E(R_p - ER_p)^2 = E\left(\sum_{i=1}^{n} w_i r_i - \sum_{i=1}^{n} w_i \mu_i\right)^2 =$

$$\sum_{i=1}^{n} w_i^2 \sigma_i^2 + 2\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_i w_j \rho_{ij} \mu_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} =$$

$$w^T C\ w$$

Modern Portfolio Theory

- The risk adjusted return R_{ADJ} is defined as

  $\alpha R_p - \sigma_p^2 = \alpha w^T R - w^T C w$

  The risk tolerance factor α expresses the relative importance of the risk and return (often set to 0.5)

- Optimization problem :

$$\max_{w} \alpha w^T R - w^T C w$$
s.t.:
$$w_i \quad 0,\ i=1,\ldots,n$$
$$\Sigma w_i = 1$$

Procedure for Portfolio Optimization

1. Estimate with historical data:
   - The individual returns $\mu_1, \ldots, \mu_n \equiv R$
   - The covariance matrix C
2. Solve
   $w_{OPT} = \max_{w} \alpha w^T R - w^T C w$
   for "optimal" portfolio weights w_{OPT}
3. Rebalance the portfolio

Repeat from 1. every n:th day.

Question: How optimal are the weights ?

How optimal are the weights

Answer: It depends on the stationarity of the price processes

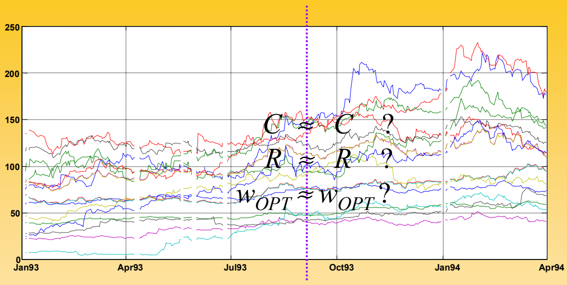
$C \approx C$?

$R \approx R$?

$w_{OPT} \approx w_{OPT}$?

Estimation of C and R and optimization | Evaluation: Computation of R_{ADJ} | *time*

How Stationary are the Processes ?



Estimation of R and C and optimization | Evaluation: Computation of R_{ADJ} | *time*

We have to PREDICT covariances C and returns R

Common methods:

1. The “Naive prediction”:
   The sample covariances and returns in a window backwards.
2. Exponentially Weighted Moving Average
3. ARCH, GARCH, D-GARCH ….

.
.
.

The Naive Prediction

At time T:
Use stock prices $Close_i(t), i=1,\ldots,n \quad t\in \{T-w,\ldots,T-1\}$
to compute sample covariances and returns:

$$\mu_j^* = \frac{1}{w}\sum_{\tau=T-w}^{T-1} Close_j(\tau)$$

$$\sigma_{ij}^* = \frac{1}{w-1}\sum_{\tau=T-w}^{T-1}(Close_i(\tau)-\mu_i^*)(Close_j(\tau)-\mu_j^*)$$

Assume that these values are valid for $t\in \{T,\ldots,T+v\}$ and optimize away

The Naive Prediction

How much past data (w) should we use ?

Empirical investigation:

w days for estimation of R and C, optimization and computation of R_{ADJ} | 20 days for estimation of R and C, optimization and computation of R_{ADJ} | *time*

For w = 20,40,60,…,300 :
Sliding window testing with two data sets with data from 1988-1997 :
24 Swedish stocks
29 American stocks

Sliding Windows

Necessary since we can’t use cross validation

time

w days for modelling | v days for evaluation

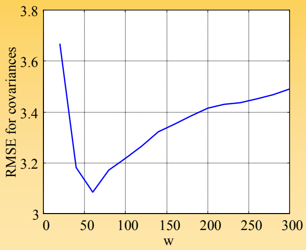
Prediction 1
Prediction 2
Prediction 2
.
. v days step
.
Prediction K

The performance is computed as the average performance for the K predictions

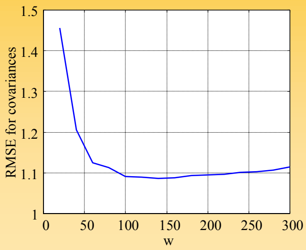
The Naive Prediction

How much past data should we use ?

Swedish stocks 88-97



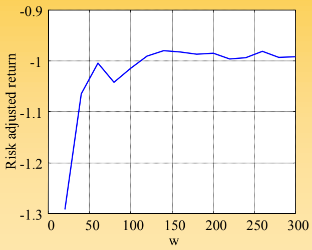
American stocks 88-97



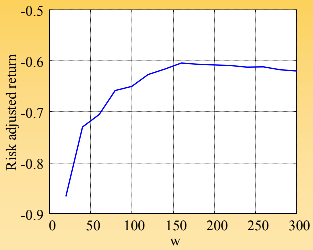
The Naive Prediction

200 days looks like a good choice

Swedish stocks 88-97



American stocks 88-97



Improving the Naive Prediction

Assumption: The Naive Prediction gains from removing certain days which can be regarded as Outliers or Noise.
The "Outlier" property for a day generalises into the future.
I.e. The same day should be removed in the next step.

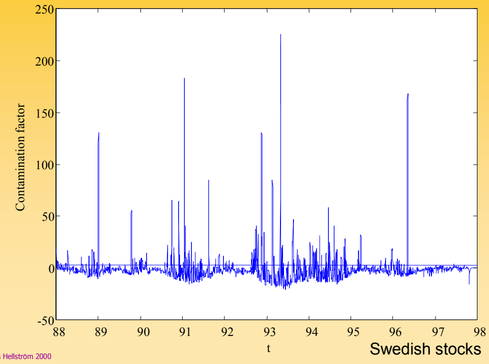
Idea:
For each step in the sliding window predictions:
    Compute a prediction where all days t with "Contamination Factor" $K_t > K_{limit}$ have been removed
    For each day t in the modelling data window:
        Remove day t, build models and predict C and R.
        Compare the prediction with one without removals.
        Update K_t depending on this comparision.
    next day t
next step

The *Remove* Algorithm:

$K_i \leftarrow 0, i = d_1, ..., d_2$
for $t = d_1 + w$ step v to $d_2 - v$
    $r \leftarrow \{j | K_j > K_{limit}, j \in (t-w, ..., t-1)\}$
    $\hat{C}(t) \leftarrow cov(t-w, t-1, r)$
    $C(t) \leftarrow cov(t, t+v-1)$
    $\hat{C}_0(t) \leftarrow cov(t-w, t-1)$
    $e_0 \leftarrow \left\| C(t) - \hat{C}_0(t) \right\|_2^2$
    for $j = t-w$ to $t-1$
        $\hat{C}_j(t) \leftarrow cov(t-w, t-1, j)$
        $e_j \leftarrow \left\| C(t) - \hat{C}_j(t) \right\|_2^2$
        $K_j \leftarrow K_j + \frac{e_0 - e_j}{e_0} 100$
    next j
next t

The Contamination Factors K_t

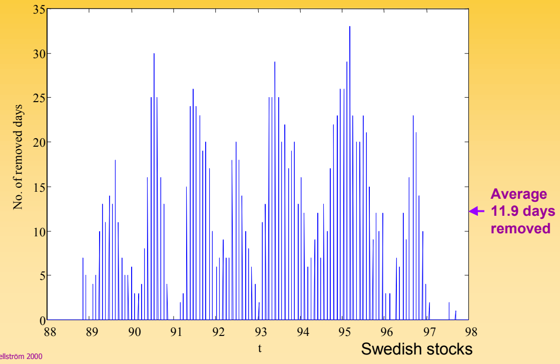
Large K_t means that day t should be removed when calculating C and R



Number of Removed Days

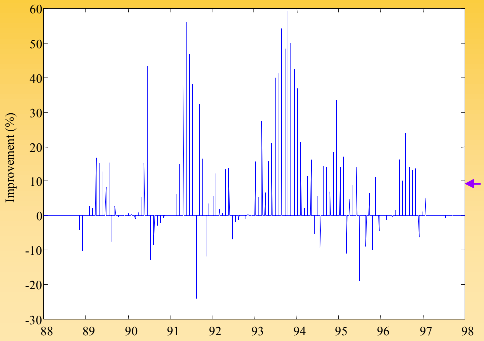
for each monthly prediction $\hat{C}(t)$.

A day j is removed iff $K_j > K_{limit}$ (K_{limit}=3 in the examples)



Improvement Relative the Naive Prediction

Improvement measured as RMSE
for the covariance predictions

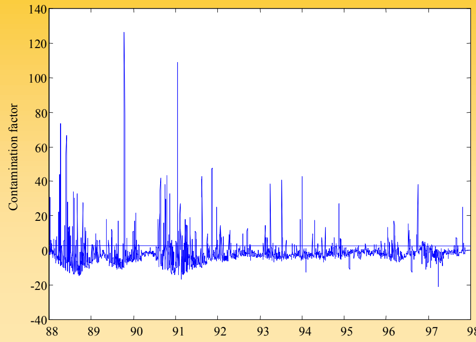


Average Improvement 8.9%

Swedish stocks

The Contamination Factors K_t

Large K_t means that day t should be removed
when calculating C and R

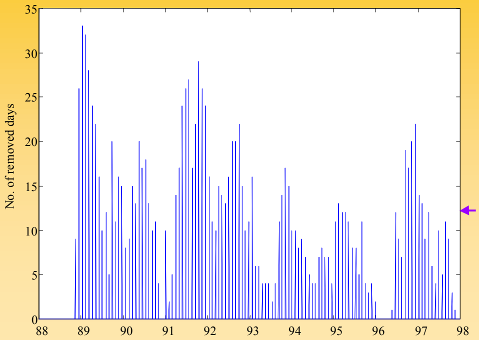


American stocks

Number of Removed Days

for each monthly prediction $\hat{C}(t)$.

A day j is removed iff $K_j > K_{limit}$ (K_{limit}=3 in the examples)

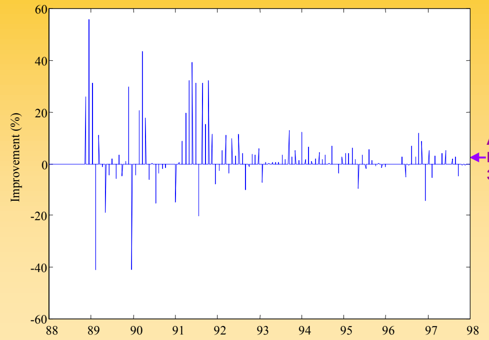


Average 11.7 days removed

American stocks

Improvement Relative the Naive Prediction

Improvement measured as RMSE
for the covariance predictions



Average Improvement 3.5%

American stocks

Results from Portfolio Optimization

Swedish stocks 88-97 :

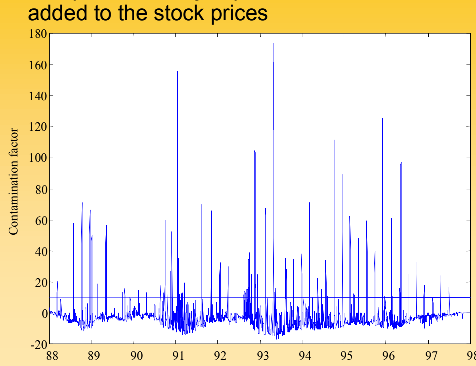
Method	R_p	σ_p^2	R_{ADJ}	Best
$\hat{C}_j$: Remove 3%	0.067	0.995	-0.959	62%
$\hat{C}_0$: Naive 200 days	0.065	1.019	-0.984	47%
Improvement	3.2%	2.4%	2.5%	
Equally balanced	0.072	1.465	-1.428	

American stocks 88-97 :

Method	R_p	σ_p^2	R_{ADJ}	Best
$\hat{C}_j$: Remove 3%	0.061	0.627	-0.597	54%
$\hat{C}_0$: Naive 200 days	0.055	0.636	-0.608	51%
Improvement	10.5%	1.4%	1.9%	
Equally balanced	0.065	0.727	-0.694	

Artificial Data

Every 50th trading day has a random noise
added to the stock prices



The noisy days are clearly detected and removed by the algorithm

Summary and Conclusions

- The RMSE for covariances is reduced by:
  8.9% for the Swedish data
  3.5% for the American data
- The increase in Risk Adjusted Return R_{ADJ} is:
  2.5% for the Swedish data
  1.9% for the American data
- Conclusion: Outliers don't affect the computation of optimal portfolios very much ?
- The *Remove* algorithm successfully detects and removes outliers in data.
- Ref: Outlier Removal for Prediction of Covariance Matrices with an Application to Portfolio Optimization
  Thomas Hellström. Technical report UMINF 00.19 Department of Computing Science, Umeå University