A Comparative Study of Covariance and Precision Matrix Estimators for Portfolio Selection

M. Senneret[1] Y. Malevergne[2,3] P. Abry[4] G. Perrin[1]
L. Jaffrès[1]

[1]Vivienne Investissement
[2]Université de Lyon - Université de Saint-Etienne
[3]EMLYON Business School
[4]CNRS, ENS Lyon

AFFI, 2013

Outline

1. Motivation
2. Setup
   - Assets & Optimizations
   - Direct Estimates
   - Factor Models
   - Shrinkage
   - Sparsity
3. Results
   - Volatility Control
   - Sharpe Control
   - Concentration Control

Estimation & Inversion

- Covariance input is **central** to Mean Variance Portfolio (MVP)
- Estimation of covariance suffers from large estimation fluctuations, even for large n (# of samples).
- Mean Variance problem $\Longrightarrow$ inversion of covariance matrix.
- Estimation error + numerical instability
  $\Longrightarrow$ MVP$\prec$ equally-weighted (EWP) - DeMiguel *et al.* (09)
- Best use of known information (var/covar) ?

Estimation & Inversion

- Covariance input is **central** to Mean Variance Portfolio (MVP)
- Estimation of covariance suffers from large estimation fluctuations, even for large n (# of samples).
- Mean Variance problem $\Longrightarrow$ inversion of covariance matrix.
- Estimation error + numerical instability
  $\Longrightarrow$ MVP$\prec$ equally-weighted (EWP) - DeMiguel *et al.* (09)
- Best use of known information (var/covar) ?

Estimation & Inversion

- Covariance input is **central** to Mean Variance Portfolio (MVP)
- Estimation of covariance suffers from large estimation fluctuations, even for large n (# of samples).
- Mean Variance problem $\Longrightarrow$ inversion of covariance matrix.
- Estimation error + numerical instability
  $\Longrightarrow$ MVP$\prec$ equally-weighted (EWP) - DeMiguel *et al.* (09)
- Best use of known information (var/covar) ?

Estimation & Inversion

- Covariance input is **central** to Mean Variance Portfolio (MVP)
- Estimation of covariance suffers from large estimation fluctuations, even for large n (# of samples).
- Mean Variance problem $\Longrightarrow$ inversion of covariance matrix.
- Estimation error + numerical instability
  $\Longrightarrow$ MVP$\prec$ equally-weighted (EWP) - DeMiguel *et al.* (09)
- Best use of known information (var/covar) ?

Estimation & Inversion

- Covariance input is **central** to Mean Variance Portfolio (MVP)
- Estimation of covariance suffers from large estimation fluctuations, even for large n (# of samples).
- Mean Variance problem $\Longrightarrow$ inversion of covariance matrix.
- Estimation error + numerical instability
  $\Longrightarrow$ MVP$\prec$ equally-weighted (EWP) - DeMiguel *et al.* (09)
- Best use of known information (var/covar) ?

Covariance or Precision?

Notations:

- Precision matrix Θ
- Covariance matrix Σ
- with $\Theta = \Sigma^{-1}$

- Θ is the input of interest
- $\hat{\Sigma}^{-1}$: poor estimator of Θ.
- Develop reliable estimators of Θ
- Compare results.

Covariance or Precision?

Notations:

- Precision matrix Θ
- Covariance matrix Σ
- with $\Theta = \Sigma^{-1}$
- Θ is the input of interest
- $\hat{\Sigma}^{-1}$: poor estimator of Θ.
- Develop reliable estimators of Θ
- Compare results.

Covariance or Precision?

Notations:

- Precision matrix Θ
- Covariance matrix Σ
- with $\Theta = \Sigma^{-1}$
- Θ is the input of interest
- $\hat{\Sigma}^{-1}$: poor estimator of Θ.
- Develop reliable estimators of Θ
- Compare results.

Covariance or Precision?

Notations:

- Precision matrix Θ
- Covariance matrix Σ
- with $\Theta = \Sigma^{-1}$

- Θ is the input of interest
- $\hat{\Sigma}^{-1}$: poor estimator of Θ.
- Develop reliable estimators of Θ
- Compare results.

Constraints as indirect regularization

- Jagannathan & Ma (2003) : introduction of constraints $\Longrightarrow$ better out-of-sample performance
- investment restrictions act as shrinkage of Σ
- How is this result related to the instability of $\hat{\Sigma}^{-1}$?

Constraints as indirect regularization

- Jagannathan & Ma (2003) : introduction of constraints $\Longrightarrow$ better out-of-sample performance
- investment restrictions act as shrinkage of Σ
- How is this result related to the instability of $\hat{\Sigma}^{-1}$?

Constraints as indirect regularization

- Jagannathan & Ma (2003) : introduction of constraints $\Longrightarrow$ better out-of-sample performance
- investment restrictions act as shrinkage of Σ
- How is this result related to the instability of $\hat{\Sigma}^{-1}$?

Outline

Assets & strategy

- Universe: $p = 211$ european stocks of the STOXX®Europe 600
- Between December 14, 2001 and January 24, 2013
- Covariance and precision estimated over 3 rolling windows: $n = 150, 200, 350$
- Every five days rebalancing
- No fees

Unconstrained Optimization

- the global MV portfolio *without restriction* w^* is given by

$$w^* = \frac{\Theta 1_p}{1_p' \Theta 1_p}$$

- Θ ? $\hat{\Sigma}^{-1}$?
- $n \leq p$?

Unconstrained Optimization

- the global MV portfolio *without restriction* w^* is given by

$$w^* = \frac{\Theta 1_p}{1_p' \Theta 1_p}$$

- Θ ? $\hat{\Sigma}^{-1}$?
- $n \leq p$?

Constrained Optimization

- *With short sales restrictions*:

$$\begin{aligned} w^* &= \operatorname{argmin}_w w'\Sigma w \\ & \quad s.t.\ 1_p' w = 1 \\ & \quad w \geq 0 \end{aligned}$$

- Lagrangian

$$L(w,\lambda,\nu) = w'\Sigma w + \lambda(1_p' w - 1) - \nu' w \ ,$$

Outline

Direct Estimates

- if $n < p$, then Σ is not full rank $\Longrightarrow$ No inversion
- if $n \simeq p$, then Σ is near criticical point $\Longrightarrow$Innaccuracy
- Use Moore Penrose Inverse
- Use diagonal matrix
- Use Identity

Direct Estimates

- if $n < p$, then Σ is not full rank $\Longrightarrow$ No inversion
- if $n \simeq p$, then Σ is near criticical point $\Longrightarrow$Innaccuracy
- Use Moore Penrose Inverse
- Use diagonal matrix
- Use Identity

Direct Estimates

- if $n < p$, then Σ is not full rank $\Longrightarrow$ No inversion
- if $n \simeq p$, then Σ is near criticical point $\Longrightarrow$Innaccuracy
- Use Moore Penrose Inverse
- Use diagonal matrix
- Use Identity

Direct Estimates

- if $n < p$, then Σ is not full rank $\Longrightarrow$ No inversion
- if $n \simeq p$, then Σ is near criticical point $\Longrightarrow$Innaccuracy
- Use Moore Penrose Inverse
- Use diagonal matrix
- Use Identity

Direct Estimates

- if $n < p$, then Σ is not full rank $\Longrightarrow$ No inversion
- if $n \simeq p$, then Σ is near criticical point $\Longrightarrow$Innaccuracy
- Use Moore Penrose Inverse
- Use diagonal matrix
- Use Identity

Outline

Factor Models

- Full rank estimate when $n < p$
- Easy and reliable estimate of Θ
- Noise Reduction
- Introduction of information via exogenous factor
- Two ways : exogenous model (Sharpe 1963, Fama French 1992), endogenous model (PCA)

Factor Models

- Full rank estimate when $n < p$
- Easy and reliable estimate of Θ
- Noise Reduction
- Introduction of information via exogenous factor
- Two ways : exogenous model (Sharpe 1963, Fama French 1992), endogenous model (PCA)

Factor Models

- Full rank estimate when $n < p$
- Easy and reliable estimate of Θ
- Noise Reduction
- Introduction of information via exogenous factor
- Two ways : exogenous model (Sharpe 1963, Fama French 1992), endogenous model (PCA)

Factor Models

- Full rank estimate when $n < p$
- Easy and reliable estimate of Θ
- Noise Reduction
- Introduction of information via exogenous factor
- Two ways : exogenous model (Sharpe 1963, Fama French 1992), endogenous model (PCA)

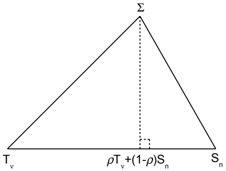
Factor Models

- Full rank estimate when $n < p$
- Easy and reliable estimate of Θ
- Noise Reduction
- Introduction of information via exogenous factor
- Two ways : exogenous model (Sharpe 1963, Fama French 1992), endogenous model (PCA)

Outline

Shrinkage Approach

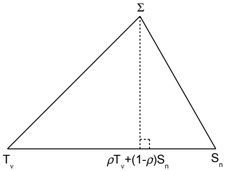
- Mix between sample matrix and model matrix, nearest to the true matrix (Ledoit and Wolf (2004)).



- Shrink with Identity or diagonal
- Shrink to precision Θ
- S_n singular $\Longrightarrow$ hard problem, moments of the Moore Penrose inverse are unknown

Shrinkage Approach

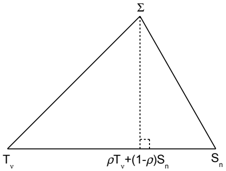
- Mix between sample matrix and model matrix, nearest to the true matrix (Ledoit and Wolf (2004)).



- Shrink with Identity or diagonal
- Shrink to precision Θ
- S_n singular $\Longrightarrow$ hard problem, moments of the Moore Penrose inverse are unknown

Shrinkage Approach

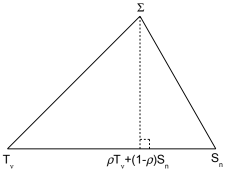
- Mix between sample matrix and model matrix, nearest to the true matrix (Ledoit and Wolf (2004)).



- Shrink with Identity or diagonal
- Shrink to precision Θ
- S_n singular $\Longrightarrow$ hard problem, moments of the Moore Penrose inverse are unknown

Shrinkage Approach

- Mix between sample matrix and model matrix, nearest to the true matrix (Ledoit and Wolf (2004)).



- Shrink with Identity or diagonal
- Shrink to precision Θ
- S_n singular $\Longrightarrow$ hard problem, moments of the Moore Penrose inverse are unknown

Outline

Sparsity Justification

- Principle of parsimony $\Longrightarrow$ Occam's razor
- *a priori* theoretical justification:
  - assets of different class are weakly related with others, e.g. bonds/stocks
  - within a class, conditional correlation may be weak
- Statistical justification: Screening effect $\Longrightarrow$ off-diagonal terms can be statistically indistinguishable from 0
- Which coefficients must be forced to 0?

Sparsity Justification

- Principle of parsimony $\Longrightarrow$ Occam's razor
- *a priori* theoretical justification:
  - assets of different class are weakly related with others, e.g. bonds/stocks
  - within a class, conditional correlation may be weak
- Statistical justification: Screening effect $\Longrightarrow$ off-diagonal terms can be statistically indistinguishable from 0
- Which coefficients must be forced to 0?

Sparsity Justification

- Principle of parsimony $\Longrightarrow$ Occam's razor
- *a priori* theoretical justification:
  - assets of different class are weakly related with others, e.g. bonds/stocks
  - within a class, conditional correlation may be weak
- Statistical justification: Screening effect $\Longrightarrow$ off-diagonal terms can be statistically indistinguishable from 0
- Which coefficients must be forced to 0?

Sparsity Justification

- Principle of parsimony $\Longrightarrow$ Occam's razor
- *a priori* theoretical justification:
  - assets of different class are weakly related with others, e.g. bonds/stocks
  - within a class, conditional correlation may be weak
- Statistical justification: Screening effect $\Longrightarrow$ off-diagonal terms can be statistically indistinguishable from 0
- Which coefficients must be forced to 0?

Principles

- Trade-off between data fidelity and promoting sparsity
- Most common methods act on precision (Boyd *et al.* 2010):

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \operatorname{Tr}(S_n\Theta) - \log\det\Theta + \lambda||\Theta||_1$$

- l_1 norm acts as a surrogate to the counting of non-zero terms
- New method acts directly on covariance, but is non convex (Bien & Tibshirani 2012)

$$\hat{\Sigma} = \underset{\Sigma}{\operatorname{argmin}} \operatorname{Tr}(S_n\Sigma^{-1}) + \log\det\Sigma + \lambda||\Sigma||_1$$

Principles

- Trade-off between data fidelity and promoting sparsity
- Most common methods act on precision (Boyd *et al.* 2010):

$$\hat{\Theta} = \underset{\Theta}{\text{argmin}}\, \mathrm{Tr}(S_n\Theta) - \log\ \det\ \Theta + \lambda||\Theta||_1$$

- l_1 norm acts as a surrogate to the counting of non-zero terms
- New method acts directly on covariance, but is non convex (Bien & Tibshirani 2012)

$$\hat{\Sigma} = \underset{\Sigma}{\text{argmin}}\, \mathrm{Tr}(S_n\Sigma^{-1}) + \log\ \det\ \Sigma + \lambda||\Sigma||_1$$

Principles

- Trade-off between data fidelity and promoting sparsity
- Most common methods act on precision (Boyd *et al.* 2010):

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \operatorname{Tr}(S_n \Theta) - \log \det \Theta + \lambda ||\Theta||_1$$

- l_1 norm acts as a surrogate to the counting of non-zero terms
- New method acts directly on covariance, but is non convex (Bien & Tibshirani 2012)

$$\hat{\Sigma} = \underset{\Sigma}{\operatorname{argmin}} \operatorname{Tr}(S_n \Sigma^{-1}) + \log \det \Sigma + \lambda ||\Sigma||_1$$

Outline

Volatility - Input: Covariance

Volatility	150	200	350	150	200	350
	Unconstrained			**Constrained**		
Basic approaches						
Sample	32.44	43.48	10.42	13.72	14.87	10.26
Identity	20.88	20.88	20.88	20.88	14.87	20.88
Diagonal	16.79	16.93	17.44	16.82	20.88	17.46
Factor models						
ACP 1F	9.03	9.16	9.93	9.88	9.93	10.43
ACP 2F	9.05	9.24	10.01	9.82	9.89	10.41
Market Model	16.43	16.60	17.17	16.47	16.65	17.2
Shrinkage						
Identity	**8.97**	9.04	9.18	9.82	9.86	**10.25**
Diagonal	9.73	**8.48**	**8.81**	**9.73**	**9.84**	**10.25**
Sparsity	16.81	16.95		13.99	14.99	

Volatility - Input: Precision

Volatility	150	200	350	150	200	350
	Unconstrained			**Constrained**		
Basic approaches						
Sample	32.44	43.48	10.42	13.72	14.87	10.26
Identity	20.88	20.88	20.88	20.88	14.87	20.88
Diagonal	16.79	16.93	17.44	16.82	20.88	17.46
Factor models						
ACP 1F	9.03	**9.16**	**9.93**	9.88	9.93	10.43
ACP 2F	9.05	9.24	10.01	**9.82**	**9.89**	**10.41**
Market Model	17.43	17.34	17.73	17.45	17.36	17.74
Shrinkage						
Identity	-	-	20.60	-	-	20.71
Diagonal	-	-	19.28	-	-	19.3
Sparsity	**7.81**	16.63	10.1	10.3	12.34	11.64

Outline

Sharpe - Input: Covariance

Sharpe	150	200	350	150	200	350
	Unconstrained			**Constrained**		
Basic approaches						
Sample	0.72	0.23	0.8	0.84	0.94	0.86
Identity	0.72	0.72	0.72	0.72	0.72	0.72
Diagonal	0.67	0.74	0.71	0.67	0.74	0.71
Factor models						
ACP 1F	0.93	1.02	0.99	0.97	1.03	0.85
ACP 2F	0.96	1.03	1.02	0.92	1.00	0.81
Market Model	0.68	0.74	0.71	0.68	0.75	0.71
Shrinkage						
Identity	**1.14**	**1.11**	**1.11**	**1.15**	**1.18**	**0.95**
Diagonal	1.05	1.08	1.09	1.05	1.09	0.86
Sparsity	0.67	0.74		0.82	0.83	

Sharpe - Input: Precision

Sharpe	150	200	350	150	200	350
	Unconstrained			**Constrained**		
Basic approaches						
Sample	0.72	0.23	0.8	0.84	0.94	0.86
Identity	0.72	0.72	0.72	0.72	0.72	0.72
Diagonal	0.67	0.74	0.71	0.67	0.74	0.71
Factor models						
ACP 1F	0.93	1.02	0.99	**0.97**	**1.03**	0.85
ACP 2F	0.96	1.03	1.02	0.92	1.00	0.81
Market Model	0.67	0.73	0.72	0.67	0.73	0.72
Shrinkage						
Identity	-	-	0.71	-	-	0.71
Diagonal	-	-	0.62	-	-	0.62
Sparsity	**1.33**	**1.05**	**1.11**	0.89	1	**0.95**

Sharpe - Input: Precision

Sharpe	150	200	350	150	200	350
	Unconstrained			**Constrained**		
Basic approaches						
Sample	0.72	0.23	0.8	0.84	0.94	0.86
Identity	0.72	0.72	0.72	0.72	0.72	0.72
Diagonal	0.67	0.74	0.71	0.67	0.74	0.71
Factor models						
ACP 1F	0.93	1.02	0.99	**0.97**	**1.03**	0.85
ACP 2F	0.96	1.03	1.02	0.92	1.00	0.81
Market Model	0.67	0.73	0.72	0.67	0.73	0.72
Shrinkage						
Identity	-	-	0.71	-	-	0.71
Diagonal	-	-	0.62	-	-	0.62
Sparsity	**1.33**	**1.05**	**1.11**	0.89	1	**0.95**

Outline

Unconstrained Optimization

	Sharpe	Turnover	Herfindahl^{-1}	Short int.
150				
Sample	0.72	27.69	1.62	-16.52
Cov-Id Shrinkage	1.14	2.71	12.51	-3.46
Sparsity Precision	1.33	0.76	14.94	-1.18
200				
Sample	0.23	23.57	0.5	-10.89
Cov-Id Shrinkage	1.11	1.05	10.07	-1.52
Sparsity Precision	1.05	0.08	48.48	-0.08
350				
Sample	0.23	23.57	0.5	-10.89
Cov-Id Shrinkage	1.11	1.77	7.98	-3.69
Sparsity Precision	1.11	0.21	30.11	-0.48

Summary

- Confirmation of the relevance of shrinkage when focusing on covariance-based optimization.
- Confirmation of the positive impact of constraints on badly estimated covariance matrix.
- Negative impact of constraints when good estimate is available.
- Precision-based optimization is sightly better than covariance when unconstrained.
- Precision-based optimization leads to much more stable portfolios.

Summary

- Confirmation of the relevance of shrinkage when focusing on covariance-based optimization.
- Confirmation of the positive impact of constraints on badly estimated covariance matrix.
- Negative impact of constraints when good estimate is available.
- Precision-based optimization is sightly better than covariance when unconstrained.
- Precision-based optimization leads to much more stable portfolios.

Summary

- Confirmation of the relevance of shrinkage when focusing on covariance-based optimization.
- Confirmation of the positive impact of constraints on badly estimated covariance matrix.
- Negative impact of constraints when good estimate is available.
- Precision-based optimization is sightly better than covariance when unconstrained.
- Precision-based optimization leads to much more stable portfolios.

Summary

- Confirmation of the relevance of shrinkage when focusing on covariance-based optimization.
- Confirmation of the positive impact of constraints on badly estimated covariance matrix.
- Negative impact of constraints when good estimate is available.
- Precision-based optimization is sightly better than covariance when unconstrained.
- Precision-based optimization leads to much more stable portfolios.

Regularization by constraint

- KKT

$$\begin{aligned} 2\Sigma w + \lambda \cdot 1_p - \nu &= 0 \\ 1' w - 1 &= 0 \\ \mathrm{Diag}(\nu)\ w &= 0 \\ w, \nu &\geq 0 \end{aligned}$$

- Equivalent to

$$2\left[\Sigma + \mathrm{Diag}(\nu) - \frac{1}{2} \cdot (1_p \nu' + \nu 1_p')\right] w + \lambda \cdot 1_p = 0$$

- Constraints impose an effective decay on correlation