Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

III OUUGUO

Recap: Eigenvalues

ovariance Matrices

Directional and tot variance

Estimation proble

aussian Models

aximum likelihood

ssues

Regularizat

Factor models

aphical models

nalized

vamnlae

xampies

References

Finance Data Science Lecture 3: Covariance Matrix Estimation

Laurent El Ghaoui

MFE 230P, Summer 2017 MFE Program Haas School of Business UC Berkeley

6/12/2017

Outline

Introduction

Motivations Recap: Eigenvalues

Covariance Matrices

Empirical covariance

Directional and total variance

Estimation problem

Gaussian Models

Maximum likelihood

Issues

Regularization

Factor models

Shrinkage models

Outlier detection

Sparse graphical models

Conditional independence
Penalized maximum-likelihood

Examples

References

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Introduction

Motivations

Recap: Eigenvalue

Empirical assertance

mpirical covariance

variance

Estimation problen

issian Models

elloe

) o audovina

Factor models

Shrinkage

diller detection

raphical models

enalized

aximum-likeliho

Outline

Introduction

Finance Data Science 3 Covariance Matrix

MFF 230P Summer 2017

Introduction

Motivations and goals

Finance Data Science 3 Covariance Matrix

Summer 2017

Motivations

MEE 230P

Covariance matrices are widely used in finance:

- Exploratory data analysis (see lecture 4).
- Risk analysis.
- Portfolio optimization.
- Outlier detection

In practice the number of data points n may be less than the number of dimensions p (assets).

This lecture: examine three estimation methods: one naïve (sample estimate), the other classical (factor model), the last modern.

Empirical covariance
Directional and total

Estimation probler

aussian Models faximum likelihood

Factor models

Outlier detection

Graphical models

Conditional independen Penalized maximum-likelihood

.

Theorem (EVD of symmetric matrices)

We can decompose any symmetric $p \times p$ matrix S as

$$S = U\Lambda U^{T} = \sum_{i=1}^{p} \lambda_{i} u_{i} u_{i}^{T},$$

where $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$, with $\lambda_1 \geq \ldots \geq \lambda_p$ the eigenvalues, and $U = [u_1, \ldots, u_p]$ is a $p \times p$ orthogonal matrix $(U^T U = I_p)$ that contains the eigenvectors u_i of S, that is:

$$Su_i = \lambda_i u_i, i = 1, \ldots, p.$$

Corollary: If S is square, symmetric:

$$\lambda_{\max}(S) = \max_{x : \|x\|_2 = 1} x^T S x.$$
 (1)

Positive semi-definite (PSD) matrices

A (square) symmetric matrix S is said to be *positive semi-definite* (PSD) if

$$\forall x, x^T S x \geq 0.$$

In this case, we write $S \succeq 0$.

From EVD theorem: for any square, symmetric matrix S:

 $S \succeq 0 \iff$ every eigenvalue of S is non-negative.

Hence we can numerically (via EVD) check positive semi-definiteness.

Finance Data Science 3. Covariance Matrix

> MFE 230P Summer 2017

Introduction Motivations

Recap: Eigenvalues

Covariance Matrices

Directional and total

Estimation problem

aussian Models

aximum likeliho

Feeter medala

Shrinkage

Outlier detection

araphical model:

Penalized

Examples

Outline

Introduction

Recap: Eigenvalues

Covariance Matrices

Empirical covariance
Directional and total variance
Estimation problem

Gaussian Model

Maximum likelihood

Regularization

Factor models
Shrinkage models

Outlier detection

Sparse graphical models

Conditional independence
Penalized maximum-likelihood
Examples

References

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Introduc

Motivations

Recap: Eigenvalue

Covariance Matrices

mpirical covariance

variance

Estimation problem

aussian Mode

Issues

Regulariza

Factor models Shrinkage

Outlier detection

raphical models

Conditional indene

enalized aximum-likelihoo

Examples

References

neierences



$$\sigma^2 = \frac{1}{m} \sum_{i=1}^m (z_i - \hat{z})^2,$$

where $\hat{z} = (1/m)(z_1 + ... + z_m)$ is the average of the z_i 's.

- How can we extend this notion to higher dimensions (with z_i's as vectors)?
- Why would we want to do that?

Note: for technical reasons the factor 1/m is often replaced with 1/(m-1), with little effect when m is large.

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivation

Recap: Eigenvalues

Covariance Matrices

Empirical covariance

Directional and total variance

Gaussian Models

faximum likelihood

egularization actor models

Outlier detection

Outlier detection

Graphical models
Conditional indepen

nalized aximum-likelihood

loforonooo

Finance Data Science
3. Covariance Matrix

MFE 230P

Summer 2017

Given a $p \times m$ data matrix $A = [a_1, \dots, a_m]$ (each row representing say a log-return time-series over m time periods), the *sample covariance matrix* is defined as the $p \times p$ matrix

$$S = \frac{1}{m} \sum_{i=1}^{m} (a_i - \hat{a})(a_i - \hat{a})^T, \ \hat{a} := \frac{1}{m} \sum_{i=1}^{m} a_i.$$

We can express S as

$$S = \frac{1}{m} A_c A_c^T,$$

where Ac is the centered data matrix:

$$A_c = (a_1 - \hat{a} \dots a_m - \hat{a})$$

Note: for technical reasons the factor 1/m is often replaced with 1/(m-1), with little effect when m is large.

Introduction

Recap: Eigenvalue:

Covariance Matrices

Empirical covariance

variance
Estimation problem

Gaussian Models

sues

Regularization

Shrinkage

Outlier detection

araphical models

Penalized maximum-likelihood

xamples

The (sample) variance along direction x is

$$\mathbf{var}(x) = \frac{1}{m} \sum_{i=1}^{m} [x^{T} (a_{i} - \hat{a})]^{2} = x^{T} S x = \frac{1}{m} ||A_{c}x||_{2}^{2}.$$

where A_c is the centered data matrix:

Hence:

- the covariance matrix gives information about variance along any direction, via the quadratic function x

 x^T Sx;
- the covariance matrix is always symmetric ($S = S^T$);
- ▶ It is also positive-semidefinite (PSD), since $x^T Sx = \mathbf{var}(x) \ge 0$ for every x.

Motivations

Recap: Eigenvalues

Covariance Matrice

Directional and total variance

Estimation problem

aussian Models Maximum likelihood

Regularization

Factor models Shrinkage

Outlier detection

Graphical models

enalized aximum-likelihood

References

Application: portfolio risk

- **Data:** Consider *n* assets with returns over one period (e.g., day) r_i , $i = 1, \dots, n$. In general not known in advance.
- **Portfolio**: described by a vector $x \in \mathbf{R}^n$, with $x_i > 0$ the proportion of a total wealth invested in asset i.
- ▶ *Portfolio return:* r^Tx ; in general not known.
- Expected return: mean value of portfolio return, given by

$$\mathbf{E} \mathbf{r}^T \mathbf{x} = \hat{\mathbf{r}}^T \mathbf{x}$$
.

with $\hat{r} = (\hat{r}_1, \dots, \hat{r}_n)$ the vector of mean returns.

Portfolio risk: Assuming return vector r is random, with mean \hat{r} and covariance matrix S, the variance of the portfolio is

$$\sigma^2(x) := \mathbf{E}_r(r^T x - \hat{r}^T x)^2 = x^T S x.$$

Finance Data Science 3 Covariance Matrix

> MEE 230P Summer 2017

Directional and total

variance

$$e_i = (0, \dots, 1, \dots, 0)$$
 (with 1 in *i*-th position, 0 otherwise).

Total variance writes:

$$\sum_{i=1}^{p} \mathbf{var}(e_i) = \sum_{i=1}^{p} e_i^T S e_i = \sum_{i=1}^{p} S_{ii} := \mathbf{Tr} \, S,$$

where the symbol ${\bf Tr}$ (trace) denotes the sum of the diagonal elements of its matrix argument.

Motivations

lecap: Eigenvalues

ovariance Matrice Empirical covariance

Directional and total variance

Estimation probler

aussian Models aximum likelihood

egularization

actor models Shrinkage

Outlier detection

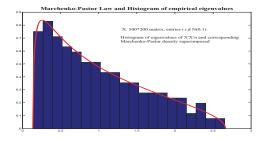
Conditional independence Penalized

amples

teferences

What is wrong with the sample covariance?

Assume we draw random data with zero mean and true covariance $S = I_p$, and look at eigenvalues of the sample estimate, when both p, n are large.



Histogram of sample eigenvalues.

- Eigenvalues should be all close to 1!
- ▶ This becomes true only when *p* is fixed and number of samples $n \to +\infty$.
- Red curve shows theoretical result from "random matrix theory" [2], which works for "large p, large n" case (see later).

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivations

Recap: Eigenvalue

impirical covariance

Estimation problem

stimation problem

aussian Model

Maximum likelihood ssues

Regularization Factor models

Shrinkage

Julier detection

Graphical models

Penalized maximum-likelihood

References

neierences



Estimation problem

In practice, the sample estimate might not work well in high dimensions; so we need to look for better estimates.

Problem: Given data points $x_1, \ldots, x_n \in \mathbf{R}^d$, find an estimate of the covariance \hat{C} .

- Many methods start with the sample estimate . . .
- ... and remove "noise" from it.

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivations

Recap: Eigenvalues

Empirical covariance

variance

Estimation problem

aussian Models Maximum likelihood

Issues

Regularization Factor models

Snrinkage

Oddier detection

Graphical models

enalized

xamples

Measuring estimation quality

Finance Data Science 3 Covariance Matrix

> MEE 230P Summer 2017

Estimation problem

Cross-validation principle:

- Remove 10 % of data points.
- Record new estimate.
- Measure average "error" between estimates.

How do we measure errors? We need a concept of distance between matrices:

- Frobenius norm (square-root of sum of squares of entries).
- If using a generative model (e.g., Gaussian), we can use Kullback-Leibler divergence (not quite a distance).

Outline

Introduction

Recap: Eigenvalues

Covariance Matrices

Empirical covariance
Directional and total variance

Gaussian Models

Maximum likelihood

Regularization

Factor models
Shrinkage models

Outlier detection

Sparse graphical models
Conditional independence
Penalized maximum-likelihoo
Examples

References

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Introduc

Motivations

necap. Eigerivatue

mpirical covariance

Directional and tota

Estimation problem

Gaussian Models

aussian Model

ximum likeliho

ssues

Regulariza

Factor models

Author dotaction

Julier detection

aphical models

onditional independe enalized

maximum-likelih

Examples

References

. 10101011000

Maximum likelihood

Let us assume that the data points are zero-mean, and follow a multi-variate Gaussian distribution: $x \simeq \dot{\mathcal{N}}(0, \Sigma)$, with Σ a $p \times p$ covariance matrix. Assume Σ is positive definite.

The Gaussian probability density function is

$$\rho(\Sigma,x) := \frac{1}{(2\pi \det \Sigma)^{\rho/2}} \exp((1/2)x^T \Sigma^{-1} x).$$

Maximum-likelihood

Finance Data Science 3. Covariance Matrix

> MFE 230P Summer 2017

Maximum likelihood

How can we find an estimate $\hat{\Sigma}$ of the true Σ , based on data points x_1, \ldots, x_n ?

Maximum-likelihood principle: maximize the likelihood

$$L(\Sigma) := \prod_{i=1}^n p(\Sigma, x_i)$$

over the variable Σ .

Maximum likelihood

Changing variables $(X := \Sigma^{-1})$, and taking the log of the likelihood, the problem can be written as

$$\max_{X} \, \log \det X - \mathbf{Tr} \, \hat{C} X$$

where \hat{C} is the sample covariance matrix. In this form, the maximum-likelihood problem is convex.

Solution: $X = \hat{C}^{-1}$, where \hat{C} is the sample covariance matrix!

Caveat: approach fails when \hat{C} is not positive-definite (e.g., when p > n!).

Issues

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Motivations

Becap: Figenval

Covariance Matric

Empirical covariano

variance

aussian Models

laximum lik

Issues

Regulariza

Factor models

Jutliar datastia

Outilor dottoolion

Graphical models

Conditional indepen

enalized aximum-likelihoo

naximum-likeliho

References

What is wrong with the sample (i.e., ML) estimate?

- Fails in (interesting) case when p > n.
- Does not handle missing data.
- High sensitivity to outliers.
- Can come up with better estimates (see next).
- Gaussian assumption is not very good with finance data.

Outline

Regularization

Finance Data Science 3 Covariance Matrix

MFF 230P Summer 2017

Regularization

Data generative model

$$v = Lz + \sigma e$$

- y is the observation (data points).
- *e* is a noise vector (assume $\mathbf{E} e = 0$, $\mathbf{E} e e^T = \sigma^2 I$).
- ightharpoonup z contains "factors" (assume **E** e = 0, **E** $ee^T = I$).
- L is a $p \times k$ loading matrix (usually, $k \ll p$).

This corresponds to a covariance matrix $\Sigma = \sigma^2 I + LL^T$.

Finance Data Science 3. Covariance Matrix

> MFE 230P Summer 2017

Motivations

Recap: Eigenvalues

Covariance Matric

Directional and total

Estimation proble

aussian Models

aximum likelihood

Regularizat

Factor models

- ...

Graphical models

onalized avimum-likelihood

xamples

Fitting factor models

Given sample covariance matrix $\hat{C} \succeq 0$, we can find L and α buy solving $\min_{\alpha \geq 0, L} \|\hat{C} - \alpha I - L L^T\|_F.$

Solution: via EVD of \hat{C} .

Finance Data Science 3. Covariance Matrix

> MFE 230P Summer 2017

Factor models

variance Estimation proble

ussian Models

Regularization

Shrinkage

Outlier detection

Conditional independ

nalized aximum-likelihood

xamples

References

In practice, we may assume that each random variable has its own noise variance.

Modified problem:

$$\min_{D,L} \|\hat{C} - D - LL^T\|_F : D \text{ diagonal, } D \succeq 0.$$

This time, no obvious solution ...

Can alternate optimization over D (easy) and L (EVD). Results in local optimum.

Computational benefits of factor models

A simple portfolio optimization problem

Risk-return trade-off:

$$\min_{x} f(x) := x^{T} C x - \lambda r^{T} x$$

- $ightharpoonup r \in \mathbf{R}^p$ (estimate) of returns.
- $x \in \mathbf{R}^p$ portfolio vector (shorting allowed).
- C (estimate of) covariance matrix.
- ▶ Parameter $\lambda > 0$ allows to choose trade-off.

The above problem is convex.

Assuming $C \succ 0$, optimal point found via $\nabla f(x) = 0$:

$$x^* = \lambda C^{-1} r$$

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Mathatian

Motivations

Covariance Matrice

Empirical covariance

variance

Estimation proble

aussian Mode

ssues

Regularizat

Factor models

utliar dataction

Outlier detection

araphical model

nalized

maximum-likelinoo Examples

zxampies

References

4□ >
4□ >
4 = >
4 = >
9 < </p>
0

Computational benefits of factor models

Direct approach

Assume $C=D+LL^T$, with $D\succ 0$, diagonal, and $F\in \mathbf{R}^{p\times k}$, with k<< p: we need to solve

$$x^* = (D + LL^T)^{-1}y$$

with $y := \lambda r$.

Direct approach: solve the $p \times p$ linear system

$$(D+LL^T)x=y$$
,

without further exploiting structure. Cost: $O(p^3)$.

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivations

Recap: Eigenvalues

Covariance Matrices

variance
Estimation problem

aussian Models

faximum likelihood sues

Regularizati

Factor models

utliar dataction

Juliler detection

raphical models

nalized

xamples

$$\begin{pmatrix} D & L \\ L^T & -I_k \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}$$

▶ Eliminate $x = D^{-1}(y - L^T z)$ and get teh $k \times k$ system in z:

$$(I + L^T D^{-1} L)z = D^{-1} Ly.$$

▶ Then solve for x via $Dx = (y - L^T z)$.

Cost: linear in p!

- ▶ Invert diagonal D: O(p).
- Form $I + L^T D^{-1} L$ and solve for z: $O(k^3 + pk^2)$.
- Get x from z: O(p).

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Methystics

Recap: Figenya

Covariance Matric

impirical covariance

variance Estimation problem

Saussian Mode

Maximum likelihoo

Regularizati

Shrinkage

Outlier detection

Conditional indene

nalized ximum-likelihood

xamples

The need for shrinkage

- ▶ Maximum-likelihood approach fails when $\hat{C} \not\succ 0$ (e.g., p > n).
- Well-conditioned estimate is often needed for subsequent use (e.g., portfolio optimization).

(Condition number of \hat{C} is $\lambda_{\max}(\hat{C})/\lambda_{\min}(\hat{C})$.)

Basic idea: Modify \hat{C} by adding a diagonal, positive-definite term.

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Motivation

Recap: Eigenvalue

Covariance Matri

Directional and tota variance

Estimation problem

aussian Models

aximum likelihood sues

Regularizatio

Factor models Shrinkage

Outlier detection

Graphical models

nalized aximum-likelihood

examples

References

110101011003



Estimate computed as a convex combination:

$$\hat{\Sigma} = \lambda I + (1 - \lambda)\hat{C},$$

where $\lambda \in (0,1)$ is a *shrinkage* factor.

- \triangleright A formula for λ is provided in [7] (has some nice statistical properties).
- Alternatively, choose λ based on cross-validation.
- Can replace the identity with another positive-definite matrix (allows to mix heterogeneous views on markets, such as news-based and price-based).
- Authors show improvements in the context of portfolio optimization.

Shrinkage



Outline

Outlier detection

Finance Data Science 3 Covariance Matrix

MFF 230P Summer 2017

Outlier detection

Outlier detection

Outlier detection problem: Consider a data point $x \in \mathbf{R}^n$. Is it very dissimilar to a data set $X = [x_1, \ldots, x_m]$?

- Arises due to errors in measurement / reporting;
- Also useful prior to running a supervised learning algorithm.
- In practice, we address the problem of ranking possible outliers in a data set (*i.e.*, we solve the above with $x = x_i$, j = 1, ..., m, and rank the dissimilarity measures.)

- First idea: evaluate the distance from the mean \hat{x}
- Issue: is the Fuclidean norm the "natural" metric to use?
- Many methods are available, including "one-class SVM", more on this later.

In what follows we assume the mean is reset to zero.

The (regularized) least-squares objective: ($\lambda > 0$ given)

$$D(x) := \min_{w,b} \|Xw - x\|_2^2 + \lambda \|w\|_2^2.$$

gives an indication of how dissimilar a point x is from the data set X:

- A small value of D(x) indicates that x can almost be expressed as a linear combination of the data points Xw, with small weights w.
- ▶ Here $\lambda > 0$ will be a parameter of the outlier detection method.

Fact: with X_c the centered data matrix, we have

$$D(x) = x^{T} (I + (1/\lambda)X_{c}X_{c}^{T})^{-1}x.$$

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivations

Recap: Eigenvalues

Covariance Matrices

Directional and total variance Estimation problem

Estimation proble

ussian Models ximum likelihood

egularization actor models

Outlier detection

raphical models

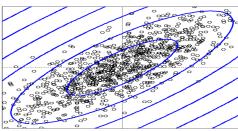
alized

xamples

Link with Mahalanobis distance

If $C \succ 0$ is a positive-definite covariance matrix, the Mahalanobis distance from a point x and a set of observations with mean \hat{x} is defined as

$$d(x) := (x - \hat{x})^T C^{-1} (x - \hat{x}).$$



The contours of the Mahalanobis distance are ellipsoids.

When $C = \hat{C} + \rho^2 I$ is a regularized estimate, with $\hat{C} = (1/m) X_c X_x^T$ a sample covariance matrix, we recover the previous distance, up to a constant factor (thus, rankings will be the same).

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Introducti

Pocan: Eigenvalues

Covariance Matrice

Empirical covariance

variance

Estimation problem

Maximum likelihood

Regularization

Factor models Shrinkage

Outlier detection

Graphical models

enalized naximum-likelihood

Examples

References

◆ロト ◆団 ▶ ◆ 豆 ▶ ◆ 豆 ・ 釣 へ ○

Outline

Introduction

Recap: Eigenvalue:

Covariance Matrices

Empirical covariance
Directional and total variance

Gaussian Models

Maximum likelihood

Regularization

Factor models
Shrinkage models

Outlier detection

Sparse graphical models

Conditional independence Penalized maximum-likelihood Examples

References

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Introducti

Motivations

Recap: Eigenvalue

Covariance Matrice:

Empirical covariance

Estimation problem

aussian Models

ussian Models vimum likelihood

Issues

Regularizat

Factor models Shrinkage

outlier detectio

Graphical models

Penalized
maximum-likelihood

Examples

References

neierences

Motivation

Finance Data Science 3 Covariance Matrix

> MEE 230P Summer 2017

Graphical models

Assume we are given prices corresponding to many assets. We'd like to draw a graph that describes the links between the prices.

- Edges in the graph should exist when some strong, natural metric of similarity exist between assets.
- For better interpretability, a sparse graph is desirable.
- Various motivations: portfolio optimization (with sparse risk term). clustering, etc.

Here we focus on exploring conditional independence within nodes.

Let us assume that the data points are zero-mean, and follow a multi-variate Gaussian distribution: $x\simeq \mathcal{N}(0,\Sigma)$, with Σ a $p\times p$ covariance matrix. Assume Σ is positive definite.

Gaussian probability density function:

$$p(x) = \frac{1}{(2\pi \det \Sigma)^{p/2}} \exp((1/2)x^T \Sigma^{-1} x).$$

where $X := \Sigma^{-1}$ is the *precision* matrix.

 $(k \neq i, j)$, the density can be factored:

where p_i , p_i depend also on the other variables.

The pair of random variables x_i , x_i are conditionally independent if, for x_k fixed

 $p(x) = p_i(x_i)p_i(x_i)$

Conditional independence

4 日 5 4 間 5 4 章 5 4 章 5 章 章

Conditional independence

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivations

Motivations

Recan: Figenvalue

ovariance Matrices
Empirical covariance

variance

aussian Model

Maximum likeli Issues

egularization

Outlier detection

San alichard ann alada

Graphical models

Conditional independence

nalized

xamples

References

The pair of random variables x_i, x_j are *conditionally independent* if, for x_k fixed $(k \neq i, j)$, the density can be factored:

$$p(x) = p_i(x_i)p_i(x_i)$$

where p_i, p_i depend also on the other variables.

Interpretation: if all the other variables are fixed then x_i, x_j are independent.

Conditional independence

Finance Data Science 3 Covariance Matrix

> MEE 230P Summer 2017

Conditional independence

The pair of random variables x_i , x_i are conditionally independent if, for x_k fixed $(k \neq i, j)$, the density can be factored:

$$p(x) = p_i(x_i)p_i(x_i)$$

where p_i , p_i depend also on the other variables.

Example: Gray hair and shoe size are independent, conditioned on age.

Theorem (C.I. for Gaussian RVs)

The variables x_i , x_j are conditionally independent if and only if the i, j element of the precision matrix is zero:

$$(\Sigma^{-1})_{ij}=0.$$

Proof.

The coefficient of $x_i x_i$ in $\log p(x)$ is $(\Sigma^{-1})_{ij}$.

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017

Motivation

Recap: Eigenvalues

ovariance Matrid

Directional and total variance

Estimation proble

ussian Models aximum likelihood

Regularizatio

Factor models Shrinkage

Jutlier detection

Graphical models

Conditional independence

Penalized maximum-likelihood

Penalized maximum-likelihood

Let us encourage sparsity of the precision matrix in the maximum-likelihood problem:

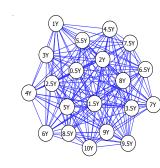
$$\max_{X}\,\log\det X - \operatorname{Tr}\,\hat{C}X - \lambda \|X\|_1,$$

with $||X||_1 := \sum_{i,j} |X_{ij}|$, and $\lambda > 0$ a parameter.

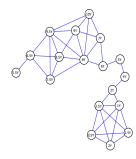
- ▶ The above provides an invertible result, even if \hat{C} is not positive-definite.
- The problem is convex.
- The result allows to discover a sparse graph revealing conditional independencies: look pairs (i, j) for which $X_{ij} = 0$.
- \triangleright Motivations for the use of the I_1 -norm: encourages sparsity.

Example

Data: Interest rates



Using covariance matrix ($\lambda = 0$).



Using $\lambda = 0.1$.

The original precision matrix is dense, but the sparse version reveals the maturity structure (an information that was not given to the algorithm).

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

Motivation

Recap: Eigenvalue

variance Matrices

Directional and total ariance

Estimation probler

Maximum likelihood

Regularization

Shrinkage

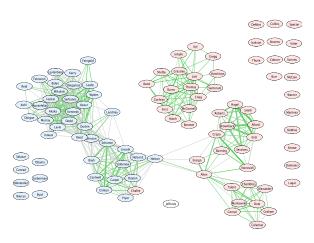
Outlier detection

Conditional independence
Penalized

Examples

Example

Data: US Senate voting, 2002-2004



Again the sparse version reveals information, here political blocks within each party.

Finance Data Science 3. Covariance Matrix

> MFE 230P Summer 2017

Examples

Outline

References

Finance Data Science 3 Covariance Matrix

> MFF 230P Summer 2017

Code

Python:

http://scikit-learn.org/stable/modules/covariance.html Implements a few methods for covariance estimation, including the sparse inverse covariance estimator.

R: http://strimmerlab.org/software/corpcor/
 Focuses on a special type of shrinkage estimator (James-Stein)

Finance Data Science 3. Covariance Matrix

> MFE 230P Summer 2017

Methystics

Recap: Eigenvalues

Covariance Matrices

empirical covariance
Directional and total
variance

Estimation problem

ussian Models ximum likelihood

egularizatio

ctor models

Outlier detection

raphical models

nditional independenc nalized iximum-likelihood

=xamples

References I

Finance Data Science
3. Covariance Matrix

MFE 230P Summer 2017



N. El Karoui.

Spectrum estimation for large dimensional covariance matrices using random matrix theory. The Annals of Statistics, 36(6):2757–2790, 2008.



N. El Karoui.

High-dimensionality effects in the markowitz problem and other quadratic programs with linear constraints: Risk underestimation.

The Annals of Statistics, 38(6):3487-3566, 2010.



Jianqing Fan, Yuan Liao, and Han Liu.

An overview of the estimation of large covariance and precision matrices.

The Econometrics Journal, 19(1):C1-C32, 2016.

Online version at https://arxiv.org/pdf/1504.02995.pdf.



Jürgen Franke, Wolfgang K. Härdle, and Christian M. Hafner.

Statistics of Financial Markets: An Introduction.

2008.

T. L. Lai and H. Xing.

Statistical models and methods for financial markets.

Springer, 2008.



Olivier Ledoit and Michael Wolf.

Honey, i shrunk the sample covariance matrix.

The Journal of Portfolio Management, 30(4):110–119, 2004.



roduction

Motivations

Recap: Eigenvalue

Covariance Matrices

inpilical covariance

variance

Estimation probler

ussian Models

rimum likelihood

Regularizati

Factor models

Shrinkage

Julier detection

raphical mode

onditional independen enalized

ximum-likelihoo

References II



Olivier Ledoit and Michael Wolf.

A well-conditioned estimator for large-dimensional covariance matrices.

Journal of Multivariate Analysis, 88:365-411, February 2004.



O.Banerjee, L. El Ghaoui, and A. d'Aspremont.

Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. Journal of Machine Learning Research, 9:485–516. March 2008.



Adam J Rothman, Elizaveta Levina, and Ji Zhu.

A new approach to cholesky-based covariance regularization in high dimensions.

Biometrika, pages 539-550, 2010.

Finance Data Science 3. Covariance Matrix

MFE 230P Summer 2017

IIIIIOddcti

Motivations

3....

nirical covariance

Directional and total

Estimation proble

aussian Models

ximum likelih

Regularizat

Shrinkage

Outlier detection

raphical models

enalized

Examples

References