Vladimir Pyrlik (talking) and *Stanislav Anatolyev*

Shrinkage for Gaussian and t Copulas in UHD

CEBA Talks
June 12, 2020

Presentation Outline

- Copulas & HD: **the briefest introduction**

- **Shrinkage for Gaussian and t Copulas in UHD**

- Further research

- **Shrinkage of large covariance matrices**: technical notes

The briefest introduction into

Copulas & HD

What are copulas?

- Copula is a multivariate CDF with $U[0, 1]$ marginals

Why are copulas important?

- **Sklar's theorem**: $F_Y(y_1, \ldots y_p) = C_{F_Y}(F_1(y_1), \ldots F_p(y_p))$
    - **The converse holds**, too: borrow a copula of one distribution, feed an arbitrary set of marginals to it, and obtain a new multivariate distribution.
- **Copulas**: a very convenient approach to generate flexible multivariate distributions.

What is "High-dimensional"?

- **HD**: number of variables p exceeds sample size n

Copulas & High Dimensionality

- **Dimensionality Curse**: in HD, either the existing frameworks are not flexible enough, or estimation is very hard to handle

- **Possible solutions**: dimensionality breakdown, structural restrictions, *shrinkage*

Shrinkage for Gaussian and t Copulas in Ultra-high Dimensions

What we do

- consider Gaussian and t copulas in (U)HD
- use shrinkage estimators of large covariance matrices
- apply the UHD t copula in large stocks portfolio allocation

Outline

1. Gaussian and t copulas: motivation and properties
2. Shrinkage estimators
3. Simulation study design and results
4. Empirical example: large portfolio allocation
5. Further research

Archimedean copulas

Pro Have analytical forms and well established properties

Pro Easily extendable to HD

Con Very rigid: parameter is always low dimensional

Pair copula constructions *aka* Vines

Pro The most flexible copula structure there can be

Con Requires structural assumptions on the data

Con HD is very hard to handle

Elliptical copulas

Pro Extendable to and remain flexible in HD

Con Traditional estimators fail in HD

- Most popular members: **Gaussian & t copulas**
- Used in a **vast variety of applications**
- **HD is appealing** (many variables or short samples)

General notation

Consider a p-dimensional r.v. $Y = (Y_1, \ldots Y_p)'$ with joint CDF $F_Y(y_1, \ldots y_p; \theta)$ and marginal CDFs $\{F_i(y_i; \theta)\}_{i=1,\ldots,p}$,

and a transformed r.v. $U = (U_1, \ldots U_p)'$, s.t. $U_i = F_i(Y_i; \theta)$.

The joint CDF $C_{F_Y}(u; \theta)$ of U is **the copula function of $\boldsymbol{F_Y}$**,

and its density $c_{F_Y}(u; \theta)$ is **the copula density function**.

Gaussian copula

$$Y \sim \mathcal{N}(\mathrm{O}_p; P), \text{ and } U_i = \Phi_{0,1}(Y_i).$$

The joint CDF of U is **the Gaussian copula**:

$$C_{\mathcal{N}}(u; P) = F_{\mathcal{N}(\mathrm{O_p}, \mathcal{P})}(\phi(u); P),$$

with the **copula density**

$$\log c_{\mathcal{N}}(u; P) = -\frac{1}{2}\log|P| - \frac{1}{2}\phi'(u) \cdot (P^{-1} - I_p) \cdot \phi(u),$$

$$\phi(u) = \left(\Phi_{0,1}^{-1}(u_1), \ldots, \Phi_{0,1}^{-1}(u_p)\right)'.$$

Gaussian copula

- **Kendall's rank correlation**: $\tau_{ij} = \frac{2}{\pi}\mathrm{asin}(P_{ij})$.

- **Spearman's rank correlation**: $\mathrm{Corr}(U) = \frac{6}{\pi}\mathrm{asin}\left(\frac{P}{2}\right)$.

- **Approx. Spearman's rank correlation**: $\mathrm{Corr}(U) \approx P$.

- **Underlying r.v. *Y* correlation**: $\mathrm{Corr}(\mathrm{Y}) = P$.

Definition & important properties

t copula

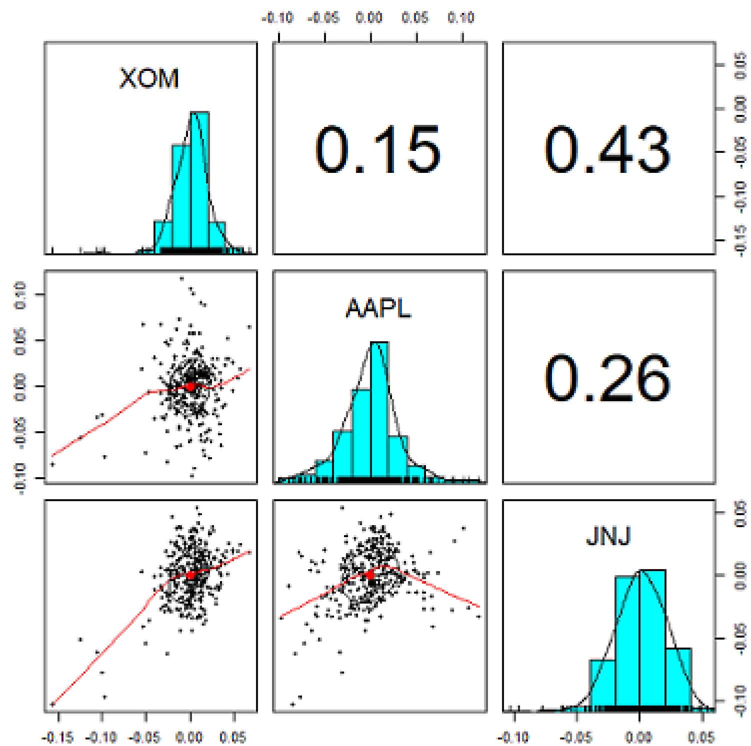
$$Y \sim MVT(P, v), \text{ and } U_i = t_v(Y_i; v).$$

Measures of dependence

- **Kendall's rank correlation**: $\tau_{ij} = \frac{2}{\pi}\operatorname{asin}(P_{ij})$.
- **Approx. Spearman's rank correlation**: $\mathrm{Corr}(U) \approx P$.
- **Underlying r.v. *Y* correlation**: $\mathrm{Corr}(\mathrm{Y}) = P$.

transformations between Y and U require knowledge of d.f.

Tail dependence



Approaches to estimation

Main idea

Given the DGP

$$X \sim F_X = C_{F_Y}(F_{X_1}(x_1; \theta_1), \ldots F_{X_p}(x_p; \theta_p); \Theta_C)$$

the copula-related part and the marginals can be estimated separately:

1. Estimate the marginals $\{\hat{F}_{X_i}(x)\}_{i=1,\ldots,p}$
2. Use the **pseudo-data** $U_i = \hat{F}_{X_i}(X_i)$ to estimate C_Y.

Approaches to estimation

Full MLE Disregards the copula structure

Pro The efficient estimator

Con Impractical even in LD, crucially impractical in HD

Pseudo-likelihood (MPLE) Treats the copula as the distribution of U.

Pro Numerically, very close to FMLE, if $\hat{F}_{X_i}$s are good.

Con Impractical in HD

Method-of-moments-like estimators

Use the measures of dependence to estimate *some of* the parameters of copula

Pro Very practical, even in HD

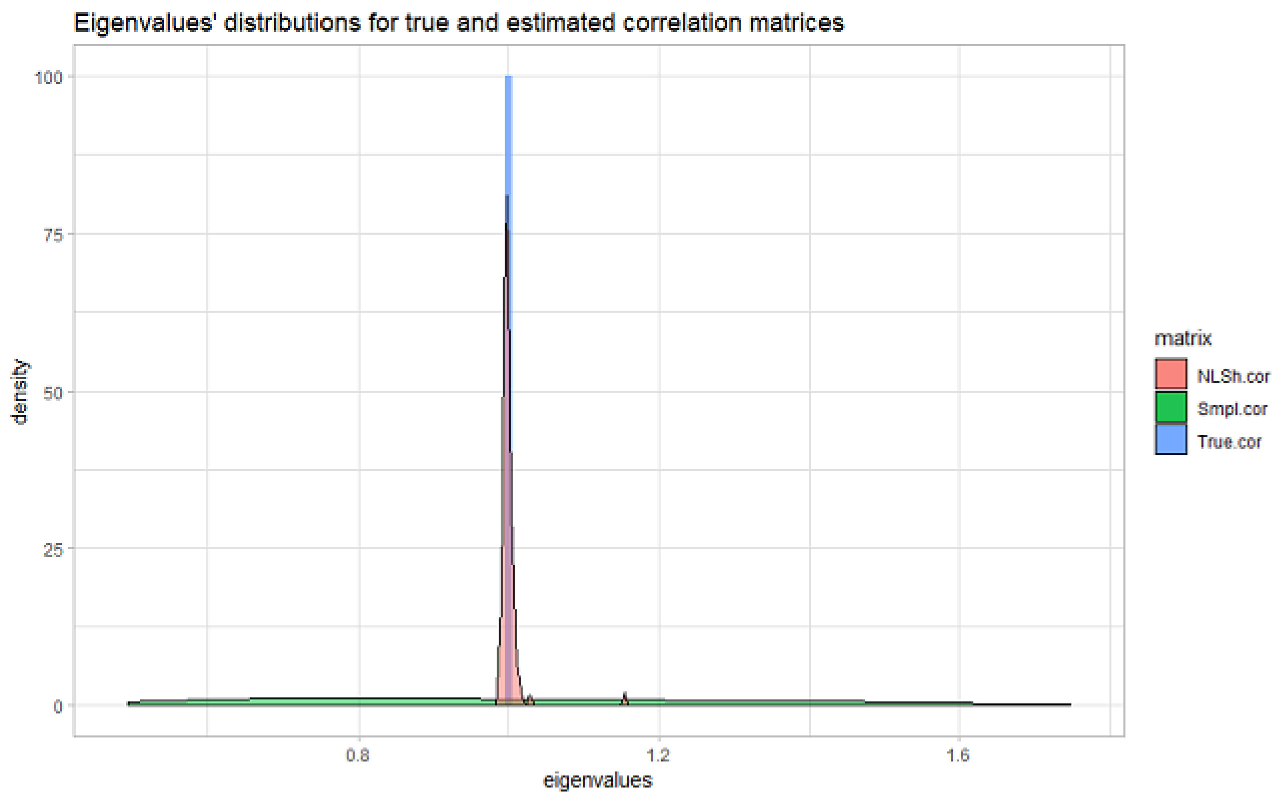
Con Not well-conditioned estimates, sometimes even in LD

Solution large covariance matrix estimators can be employed

MPLE needs to be used to estimate the remaining parameter of t copula

The intuition of

Large covariance matrices shrinkage estimators



The variety of

Large covariance matrices shrinkage estimators

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management, 30(4), 110-119.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis, 88(2), 365-411.

Ledoit, O., & Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. The Annals of Statistics, 40(2), 1024-1060.

Ledoit, O., & Wolf, M. (2017). Numerical implementation of the QuEST function. Computational Statistics & Data Analysis, 115, 199-223.

Ledoit, O., & Wolf, M. (2018). Analytical nonlinear shrinkage of large-dimensional covariance matrices. University of Zurich, Department of Economics, Working Paper, (264).

The simulation study

Given p, n, P, $\{F_{X_i}\}_{i=1,...,p}$, $C_Y(u; P)$ we find and compare $\hat{P}$s:

1. Simulate data X from $C_Y(\{F_{X_i}(x)\}_i; P)$ and switch to the pseudo-data $U_i = \hat{F}_{X_i}(X_i)$.
2. Estimate P from a variety of $\mathrm{Corr}(U)$ related measures.
3. Evaluate the quality of estimates.
4. Repeat 1 - 3 sufficient number of times.

Some details of

The simulation study

The DGP and data

- **Copulas**: Gaussian and t
- **Dimensions**: $p \in \{10, 100, 1000\}$
- **Matrix parameter**: identity or a non-sparse correlation
- **Dimensionality**: $\frac{1}{20} \leq \frac{p}{n} \leq 20$ **UHD!**

Other: skew-t marginals, d.f.=8 for t copula, B>1000

Some details of

The simulation study

The traditional estimators

- Sample analog of Kendall's rank correlation
- Sample analog of approx. Spearman's rank correlation

Suggested alternative estimators

- Linear and non-linear shrinkage estimators of pseudo-observations correlation

for t copula, the d.f. parameter is estimated via MPLE

The simulation study

Estimates quality criteria

- **Sanity check**: positive-definiteness of $\hat{P}$
- **Closeness of the parameters values**: Euclidean loss,

$$EL(P, \hat{P}) = ||\text{vech}(P - \hat{P})||_E^2$$

- **Closeness of the copulas**: Kullback - Leibler IC,

$$KLIC_{C_{\hat{P}}|C_P} = \text{E}_{C_P}[\log c_P(u) - \log c_{\hat{P}}(u)]$$

Main results

Well-conditioned estimates

- The traditional estimators are only positive-definite under LD.
- Shrinkage-based estimators guarantee positive-definiteness.

Main results

Distance measures

- Shrinkage-based estimators generally outperform the traditional ones, very significantly in HD
- Non-linear shrinkage tends to outperform the linear one
- Few settings where the linear shrinkage is better
  - Not very dispersed true eigenvalues
  - Sparse (identity) correlation matrix & short samples

even then the gain of LSh is minimal

Main results

A typical performance slice #1

The "winning" KLIC: p=100, arbitrary P, Gaussian copula

Main results

A typical performance slice #2

The "winning" KLIC: p=1000, identity P, t copula

UHD t copula application for

Large portfolio allocation

Empirical example

What we do

We apply t copula shrinkage estimation to allocate large portfolios of stocks and compare their performance with portfolio choices based on the traditional estimators.

Outline

- **Copulas & portfolios**: motivation and intuition
- **Why UHD?** Data description
- **Modeling technique**
- **Main results and discussion**

Copulas & portfolios

- Portfolio allocation is **one of the most popular applications** for the models of joint distribution of assets prices
- Portfolios' performance crucially depends on the **ability of the models to capture essential properties** of the data
  - Copulas allow to create **flexible models**
  - **Tail dependence** proved to be of dramatic importance

Why UHD?

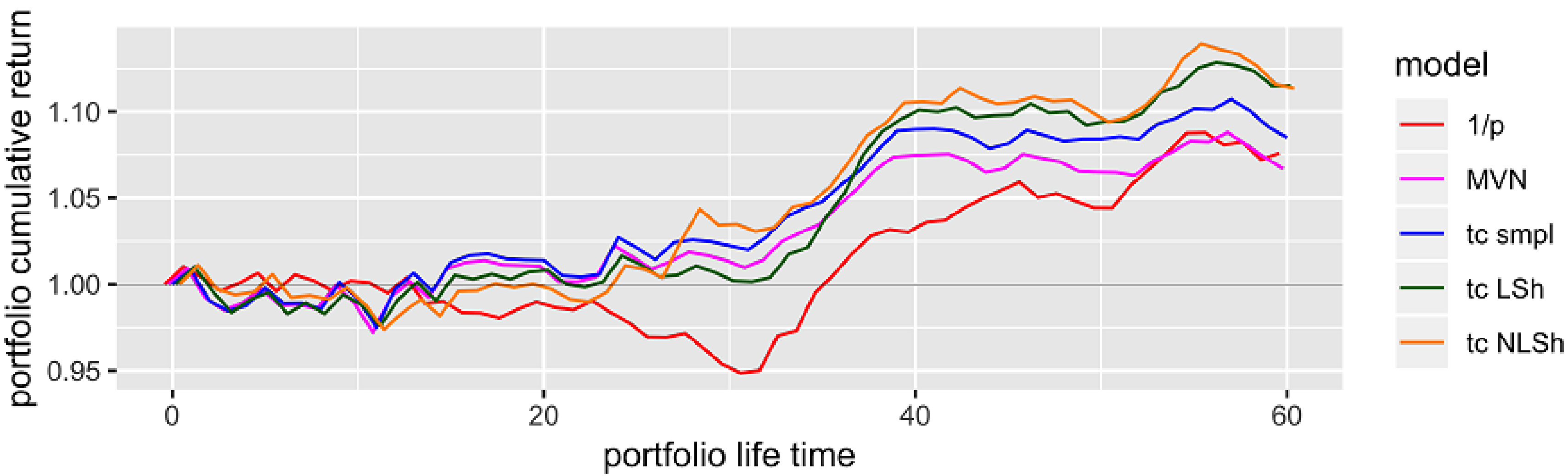
- We use the listing of Wilshire 5000 index and obtain data on 3600+ assets from CRSP
- Each portfolio contains **p=2500 assets** chosen randomly
- The univariate models of returns dynamics are simple
- Only 6 months of 2017 are used for estimation, **n=120 obs**
- Thus, the dimensionality is **$p/n > 20$** **UHD!**

The timeline

- January - June, 2017: **training period**
  4 models of assets prices joint distr. are estimated
- end of June, 2017: **portfolio allocation**
  each model produces Sharpe-maximizing portfolio
- July - September, 2017: **portfolio lifetime**
  actual portfolio value is observed
- end of September, 2017: **performance measurement**
  the final return and maximum drop-down are tracked

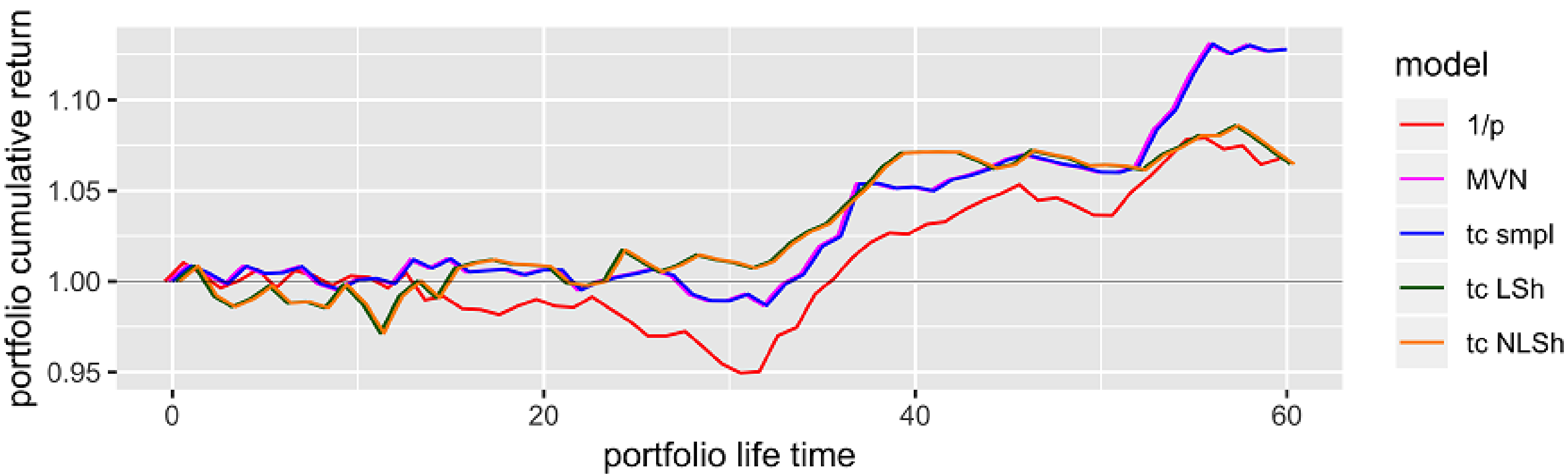
Main results

Lifetime portfolios value dynamic: Example #1



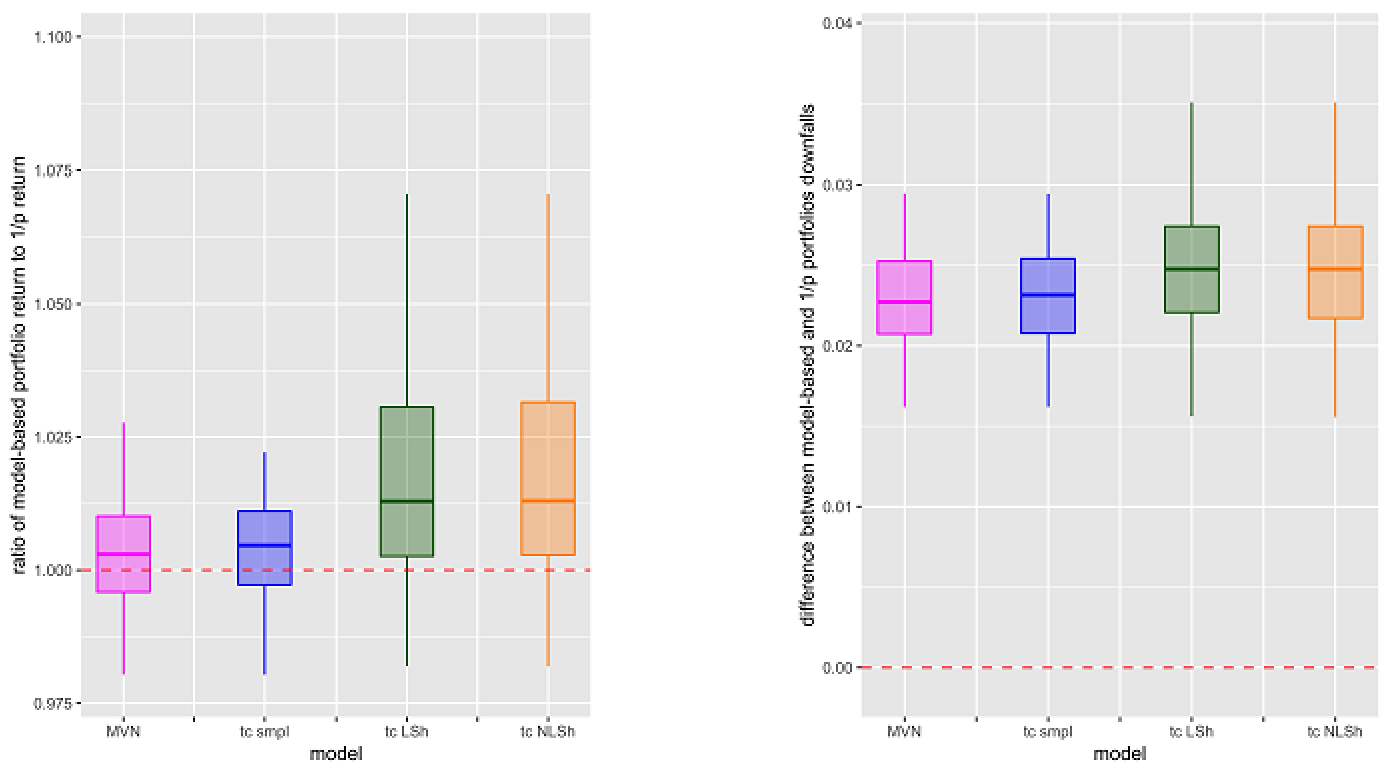
Main results

Lifetime portfolios value dynamic. Example #2



Main results

Portfolios performance overview



Discussion and Further Research

Results & further research

1. Shrinkage is a powerful tool of estimation for copulas

Further in this paper: other estimators of matrix parameters

Another work: extended copulas (skewed versions)

2. HD Copula modeling is beneficial in portfolio allocation

Another work: a profound dynamic model of joint distribution of assets prices based on copula and shrinkage

Another work: copula-based forecast combination technique

Technical notes on

Shrinkage estimation

of large covariance matrices

Large covariance matrix estimation problem

Assume $x = (x_1, ... x_p)$, $Ex = O_p$, $E[x'x] = \Sigma$ p.d.

The problem then is to estimate Σ, given data $\{X_i\}_{i=1,...n} \sim$ iid.

$S_n \equiv n^{-1}X'X$ performs well, when "p is small".

For "large p", S_n is neither consistent, nor well-conditioned.

Large *p* assumes $p \geq n$, normally even $p >> n$.

Statistical way of shrinkage

$$S_n^* = \rho_1 D + \rho_2 S_n$$

The idea

Deviate from S_n to D s.t. S_n^* is close to Σ & well conditioned.

Alternative versions

- **Choice of $\boldsymbol{D}$**: identity, equicorrelated estimate of Σ, etc.
- **Approaches to choose $\boldsymbol{\rho}$s**: heuristically, relate to p/n, *optimize*

Random Matrix Theory Approach

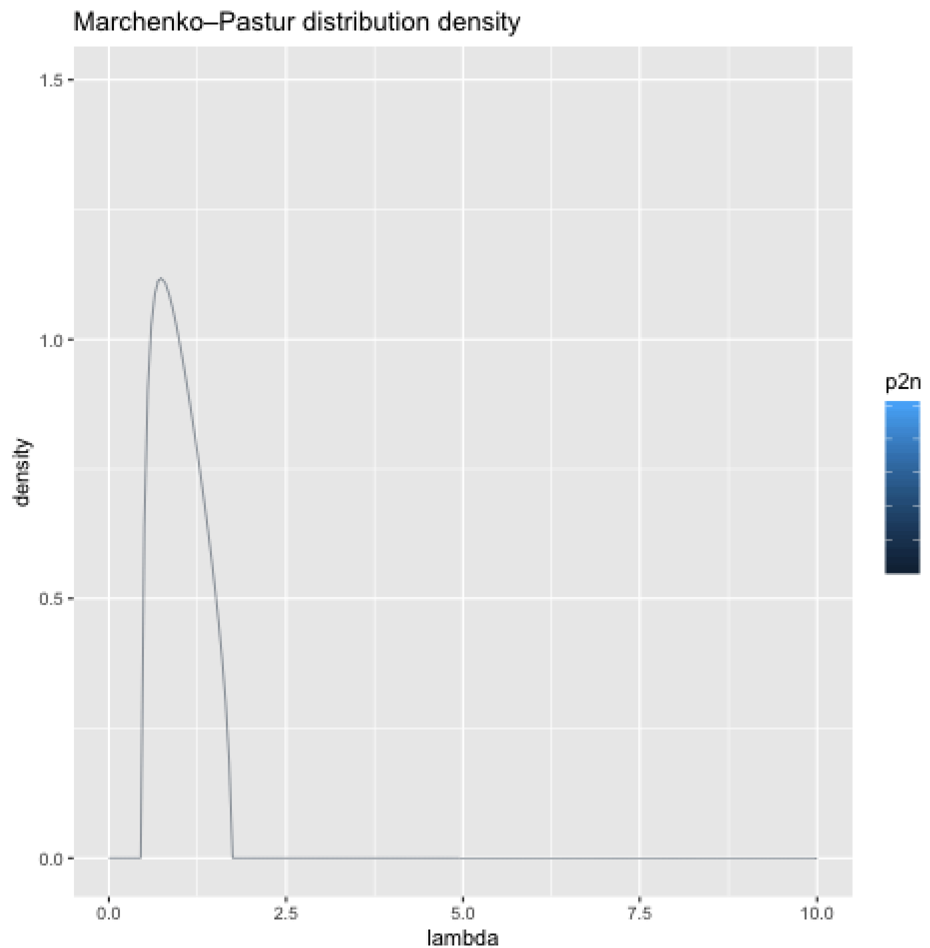
Marchenko-Pastur Law

To the same setup, $X_i \sim$ iid, $\mathrm{E}X_i = \mathrm{O}_p$, $\mathrm{E}[X_i^{'}X_i] = \Sigma$, add:

- $\{\lambda_1, \ldots \lambda_p\}$ are Σ's eigenvalues
- $\{l_1, \ldots l_p\}$ are the eigenvalues of S_n, given p, n
- **general asymptotics setup**:
  $p \longrightarrow \infty, n \longrightarrow \infty, \frac{p}{n} \longrightarrow c \in (0, \infty)/\{1, \infty\}.$

The limiting distribution of ls is **Marchenko-Pastur distribution**

Marchenko-Pastur distribution the cont. part



Eigenvalues of large S_n vs Σ

- The distributions of λs, ls and the limiting MPD share the same **grand mean**

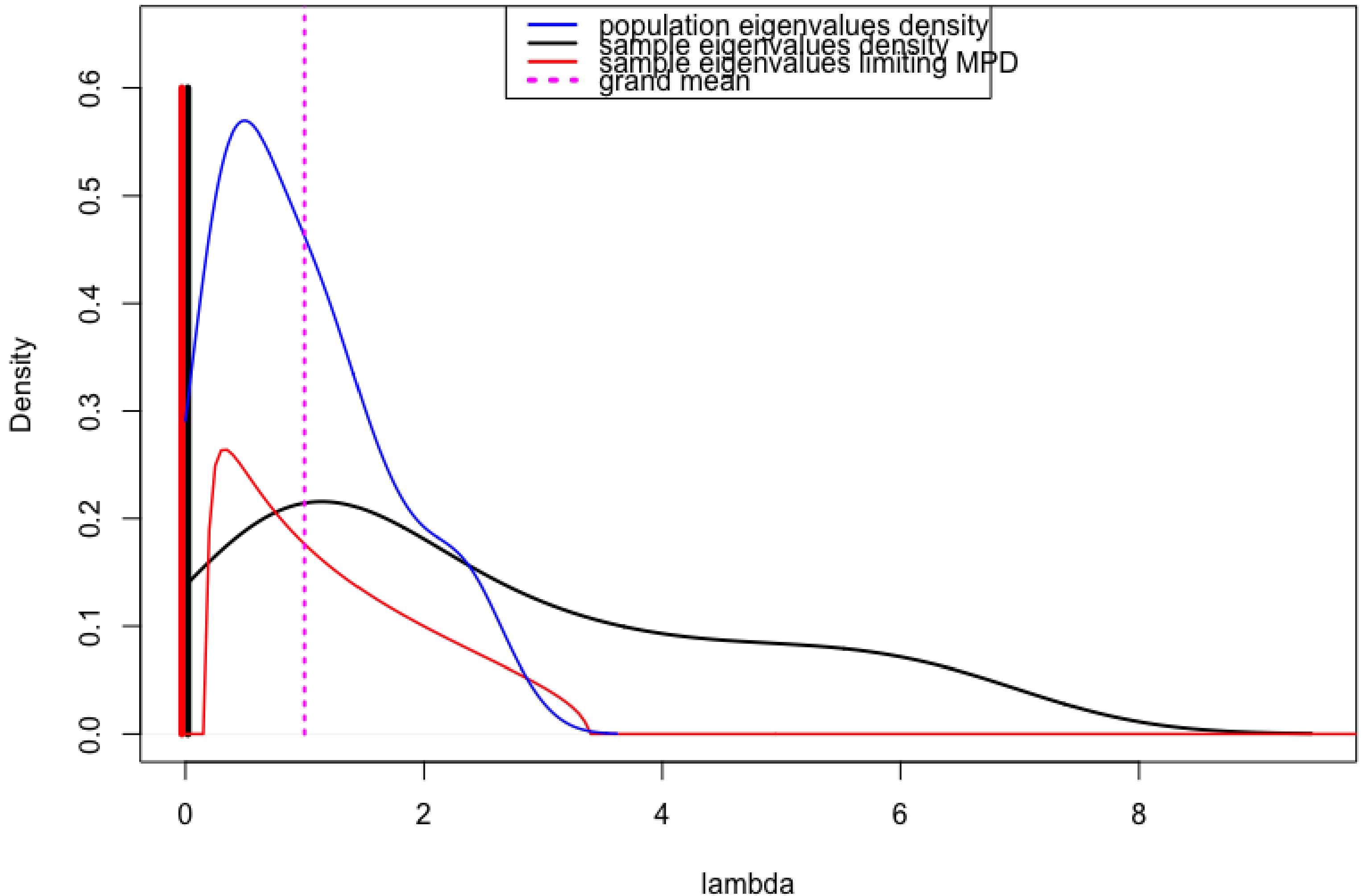
- Given particular $p, n, \; p > n$, sample distribution of ls is "overdispersed"

- The smaller e.v. are biased downwards, and the larger e.v. are biased upwards

- The limiting distribution of ls is "overdispersed", too.

- But it can be used to estimate the "range" of the population e.v.

Population and sample eigenvalues distributions, n=20, p=50



Eigenvalues shrinkage

The idea

If we "knew" the limiting MPD of ls, we could use it to shrink the sample ls towards λs.

Assume, such shrunk ls are available:

$$l^* = \rho_1 \mu + \rho_2 l.$$

Then, given $S_n = \Gamma_n \mathrm{diag}\{l_i\}_{i=1,...,p} \Gamma_n^{'}$, we can change S_n to

$$S^* = \Gamma_n \mathrm{diag}\{l_i^*\}_{i=1,...,p} \Gamma_n^{'},$$

that is potentially a better estimate of Σ.

Structural & RMT Shrinkage

Structural shrinkage:

$$S_n^* = \rho_1 D + \rho_2 S_n$$

RMT Shrinkage:

$$S^* = \Gamma_n \mathrm{diag}\{\rho_1 \mu + \rho_2 l_i\}_{i=1,...,p} \Gamma_n^{'}$$

Set $D = I_p$, and $\rho_{1,2}$ independent of ls, **these are equivalent**.

Constant shrinkage intensity estimator

"Honey, I shrunk the sample covariance matrix"

The optimal linear shrinkage estimator of Ledoit & Wolf (2004)

Ledoit, O., & Wolf, M. (2004). Honey, I shrunk the sample covariance matrix. The Journal of Portfolio Management, 30(4), 110-119.

Ledoit, O., & Wolf, M. (2004). A well-conditioned estimator for large-dimensional covariance matrices. Journal of multivariate analysis, 88(2), 365-411.

Optimal Linear Shrinkage

Optimal linear shrnkage of Ledoit & Wolf (2004) is a *constant shrinkage intensity estimator* S^{**} such that

$$S^{**} = \text{argmin}_{S^*(\rho_1,\rho_2)=\rho_1 I_p+\rho_2 S_n} ||\Sigma - S^*(\rho_1, \rho_2)||^2.$$

The idea

Since the distribution of the sample eigenvalues is overdispersed around the population eigenvalues distribution, let's relate this overdispersion to the difference between the population and sample covariance matrices to estimate the optimal shrinkage.

Optimal Linear Shrinkage

Eigenvalues overdispersion

$$p^{-1}\mathrm{E}\left[\sum_{i=1}^{p}(\lambda_i - l_i)^2\right] = p^{-1}\sum_{i=1}^{p}(\lambda_i - \mu)^2 + \mathrm{E}||S_n - \Sigma||^2.$$

The resulting estimator

$$S_n^{**} = \frac{b_n^2}{d_n^2} m_n I_p + \frac{d_n^2 - b_n^2}{d_n^2} S_n$$

Towards nonlinear shrinkage

Undershrinkage

The optimal linear shrinkage estimator is optimal, but it's linear.

Observation: it might produce "undershrinkage".

Potential improvement: make $l^{**}(l)$ a nonlinear function.

Particularly, relate *shrinkage intensity* to the sample eigenvalues:

$$l_i^{**}(l_i) = \rho_1(l_i)\mu + \rho_2(l_i)l_i.$$

Derivation notes on

Optimal Linear Shrinkage

OLSh: the setup

- $\{X_i\}_{i=1,...n} \sim \text{iid}(\text{O}_p, \Sigma)$
- $S_n \equiv n^{-1} X' X$
- $S(\rho, v) \equiv \underbrace{\rho v}_{\equiv \rho_1} \, I_p + \underbrace{(1 - \rho)}_{\equiv \rho_2} S_n$

OLSh: the setup

- **the optimization problem**:

$$(\hat{\rho}, \hat{v}) = \underset{\rho, v}{\arg\min}\ \mathrm{E}\left[\left|\left| S(\rho, v) - \Sigma \right|\right|^2 \right],$$

$$\Sigma^* \equiv S(\hat{\rho}, \hat{v}).$$

OLSh: the infeasible estimator

- **the solution to the optimization problem**:

$$\Sigma^* = \frac{\beta^2}{\delta^2}\mu I_p + \frac{\alpha^2}{\delta^2}S_n,$$

- where:

$$\mu = p^{-1}\mathrm{trace}(\Sigma),\ \alpha^2 = ||\Sigma - \mu I_p||^2,\ \beta^2 = \mathrm{E}\left[||S_n - \Sigma||^2\right],$$

$$\delta^2 = \alpha^2 + \beta^2 = \mathrm{E}\left[||S_n - \mu I_p||^2\right].$$

- the resulting estimator Σ^* is infeasible

OLSh: the infeasible estimator

$$\mathrm{E}\left[||S(\rho, v) - \Sigma||^2\right] = \rho^2||\Sigma - vI_p||^2 + (1-\rho)^2\mathrm{E}\left[||S_n - \Sigma||^2\right]$$

- the first term spits out $\hat{v}$:

$$||\Sigma - vI_p||^2 = ||\Sigma||^2 - 2v\mathrm{trace}(\Sigma) + v^2,$$

$$\hat{v} = p^{-1}\mathrm{trace}(\Sigma) \equiv \mu.$$

- then $\hat{\rho} = \frac{\mathrm{E}\left[||S_n - \Sigma||^2\right]}{\mathrm{E}\left[||S_n - \mu I_p||^2\right]} = \frac{\beta^2}{\delta^2}$ – **shrinkage intensity**

OLSh: why eigenvalues?

- $\mu = p^{-1}\sum_{i=1}^{p}\lambda_i = \mathrm{E}\left[p^{-1}\sum_{i=1}^{p} l_i\right]$
- then $\alpha^2 + \beta^2 = \delta^2$ becomes

$$p^{-1}\mathrm{E}\left[\sum_{i=1}^{p}(l_i - \mu)^2\right] = p^{-1}\sum_{i=1}^{p}(\lambda_i - \mu)^2 + \mathrm{E}\left[||S_n - \Sigma||^2\right]$$

- and the eigenvalues of Σ^* are $\lambda_i^* = \frac{\beta^2}{\delta^2}\mu + \frac{\alpha^2}{\delta^2} l_i$.
- this guarantees that Σ^* is p.d.

OLSh: feasible & asymptotically optimal estimator

the idea

Find sample analogs of $\mu, \ \alpha, \ \beta$ s.t. under *general asymptotics* the properties of the feasible estimator are the same as of Σ^*.

the solution

- $m_n \equiv \text{trace}(S_n)$
- $d_n \equiv ||S_n - m_n I_p||^2$
- $\bar{b}_n \equiv n^{-2}\sum_{k=1}^{n} ||X_k' X_k - S_n||^2$, and $b_n \equiv \min(\bar{b}_n, d_n)$
- $a_n \equiv d_n - b_n$.

Thank you!