Econometrics Reading Group

Shrinkage in Large Covariance Matrices Estimation

Vladimir Pyrlik

November 15, 2019

Hi! Thanks for watching :)

This presentation is an R-markdown based html-document, created in Xaringan.

- The web version is available at https://vpyrlik.github.io/erg1511/slides.html

NB! Before viewing the slides, give it some time to load, o.w. some elements (math!) won't look nice. If it happened, reload the page and wait a bit.

- The pdf version is available at https://vpyrlik.github.io/erg1511/slides.pdf

- The animated picture from slide 10 is available at https://vpyrlik.github.io/erg1511/stuff/mpd.gif

- Questions? Feel free to email me **vladimir.pyrlik@cerge-ei.cz**

Outline

1. Setup

Covariance matrix estimation problem. What is "Large"?

2. Shrinkage: Statistics vs RMT

Two alternative views of shrinkage & and how they relate to each other

3. Recent Shrinkage Estimators

Optimal linear & nonlinear shrinkage

Setup

Covariance matrix estimation problem

What is "Large"?

Large covariance matrix estimation problem

Assume $x = (x_1, \ldots x_p)$, $\mathbb{E}x = \mathbb{O}_p$, $\mathbb{E}[x'x] = \Sigma$ p.d.

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

Bring some structure into S_n

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

The idea

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

Alternative versions

- **Choice of D**: identity, equicorrelated estimate of Σ, etc.
- **Approaches to choose ρs**: heuristically, relate to p/n, *optimize*

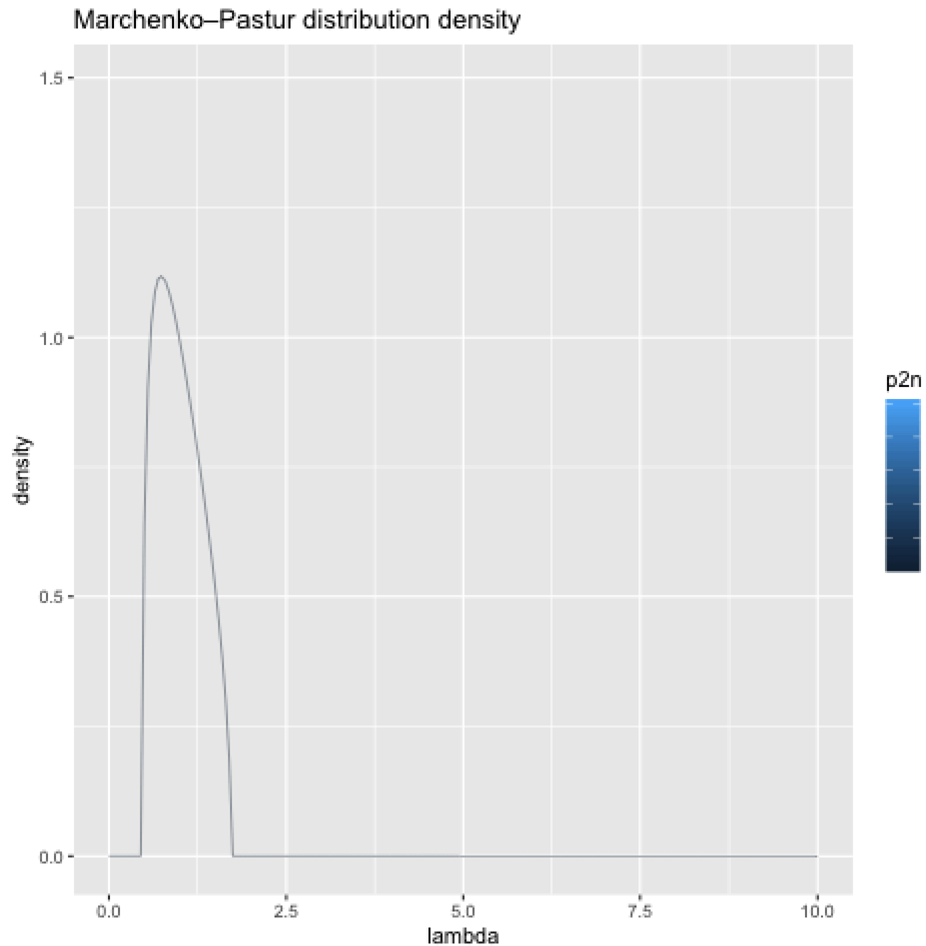
Random Matrix Theory Approach

Marchenko-Pastur Law

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

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

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


Marchenko–Pastur distribution density
density
lambda
p2n
0.0
0.5
1.0
1.5
0.0
2.5
5.0
7.5
10.0
1
2
3
4
5

Eigenvalues of large S_n vs Σ

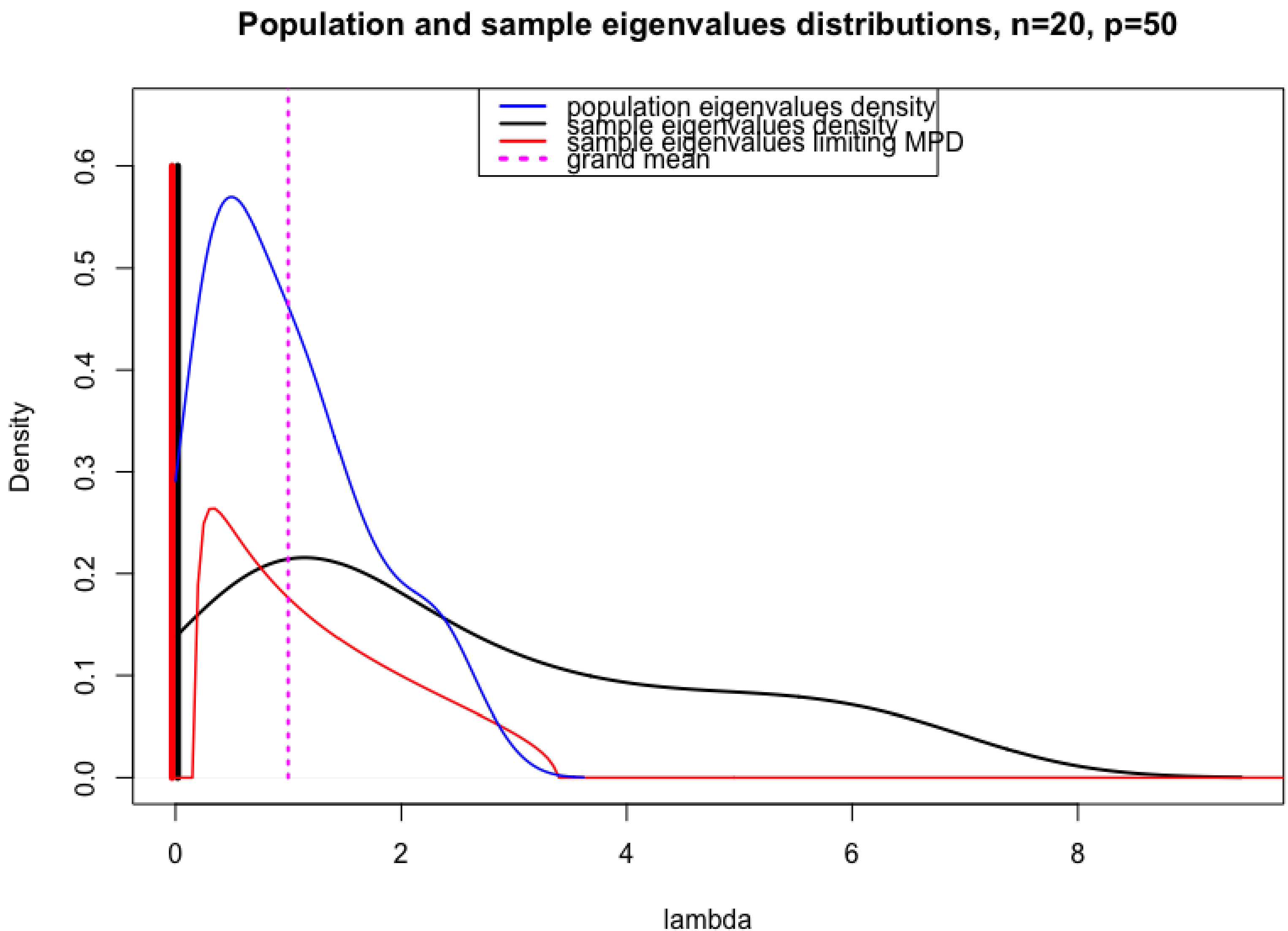
- The distributions of λs, ls and the limiting MPD share the same **grand mean**

- 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
population eigenvalues density
sample eigenvalues density
sample eigenvalues limiting MPD
grand mean
Density
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
2
4
6
8
lambda

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

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

that is potentially a better estimate of Σ.

Structural & RMT Shrinkage

Structural shrinkage:

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

RMT Shrinkage:

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

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

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

The resulting estimator

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

is asymptotically optimal in terms of Frobenius error
and consistent under general asymptotics

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:

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

Nonlinear shrinkage estimation

Oracle estimator of Ledoit & Péché (2011) and

Optimal nonlinear shrinkage of Ledoit & Wolf (2012)

Ledoit, O., & Péché, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. Probability Theory and Related Fields, 151(1-2), 233-264.

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.

Do we still have time?

Or energy?

Optimal Linear Shrinkage

derivation notes

OLSh: the setup

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

$$(\hat{\rho}, \hat{\nu}) = \arg\min_{\rho,\nu} \mathbb{E}\left[||S(\rho,\nu) - \Sigma||^2\right],$$

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

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:

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

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

the resulting estimator Σ^* is infeasible

but it can be estimated :D

OLSh: the infeasible estimator

derivation sketch

$$\mathbb{E}\left[||S(\rho,\nu)-\Sigma||^2\right] = \mathbb{E}\left[||\rho\nu I_p + (1-\rho)S_n - \Sigma||^2\right] = \dots$$

$$\dots = \rho^2||\Sigma - \nu I_p||^2 + (1-\rho)^2\mathbb{E}\left[||S_n - \Sigma||^2\right]$$

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

$$||\Sigma - \nu I_p||^2 = ||\Sigma||^2 - 2\nu\text{trace}(\Sigma) + \nu^2,$$

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

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

OLSh: why eigenvalues?

shifting to eigenvalues

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

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

- 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!