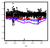
Elliptical Distributions in High Dimensions

Wolfgang Karl Härdle
Ostap Okhrin
Irina Pimenova

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

S&P500

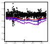
- capitalization-weighted index based on the common stock prices of 500 American companies



Challenges

Distribution and Dependency

- Risk management
  - probability of extreme events
  - VaR
- asset pricing
- asset allocation



Semi-parametrics

- 500×500 covariance matrix; short time series → questionable estimates
- 500 dimensions → curse of dimensionality

"There's always an element of risk. No one has a crystal ball. OK, I have one, but no one knows how it works."

Outline

1. Motivation ✓
2. Semi-parametrics
3. Covariance matrix estimation
4. Application

Semi-parametrics in Ellipsoids

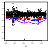
- Elliptical p.d.f.:

$$f_Y(y) = |\Sigma|^{-1/2} g\{(y-\mu)^\top \Sigma^{-1}(y-\mu)\}$$

  - y - returns
  - μ - mean
  - Σ - covariance

- **Example:** normality $g(z) = \frac{1}{(pi)^{-p/2}} \exp(-z/2)$

- **Idea:** Estimate $g_\Sigma \;\rightarrow\; f_Y(y)$



More on Ellipsoids

- If Y has an elliptical distribution, it can be represented

$$Y = \mu + RA^\top \mathcal{U}$$

with $\mathcal{U} \sim U$ on a sphere $\{t \in \mathbb{R}^p : ||t^\top t|| = 1\}$

- Useful property: $(Y-\mu)^\top \Sigma^{-1}(Y-\mu) \stackrel{\mathcal{L}}{=} R^2$
- P.d.f. of R:

$$g_R(r) = 2s_d r^{p-1} g(r^2) \text{ with } s_d = \pi^{p/2}\Gamma^{-1}(p/2)$$

More on Ellipsoids

$$g_R(r) = 2s_d r^{p-1} g(r^2) \text{ with } s_d = \frac{\pi^{p/2}}{\Gamma(p/2)}$$

- P.d.f. of R transformed into p.d.f. of R^2:

$$g_{R^2}(r) = \frac{1}{2\sqrt{r}} g(r)\sqrt{r} = s_d r^{p/2-1} g(r)$$

- Employ estimability of $g_{R^2}(r)$

$$g(r) = s_d^{-1} r^{1-p/2} g_{R^2}(r)$$

- Liebscher-transformed

$$g(r) = s_d^{-1} r^{1-p/2} \psi'(r) h\{\psi(r)\}$$

  - h - p.d.f. of $\psi\{(Y-\mu)^\top \Sigma^{-1} (Y-\mu)\}$

Density estimation

1. Estimate covariance matrix : $\widehat{\Sigma}_n$
2. Get non-parametric kernel density estimate h of $\xi_i = \psi\{(Y-\mu)^\top \Sigma^{-1}(Y-\mu)\}$

$$\widehat{h}_n(x, \omega_n; \widehat{\Sigma}_n) = \frac{1}{n\omega_n} \sum_{i=1}^{n} [\kappa\{(x - \widehat{\xi}_i)\omega_n^{-1}\} + \kappa\{(x + \widehat{\xi}_i)\omega_n^{-1}\}]$$

3. Get estimate of $\widehat{g}$

$$\widehat{g}_n(r; \widehat{\Sigma}_n) = s_d^{-1} r^{-p/2+1} \psi'(r) \widehat{h}_n(x, \omega_n; \widehat{\Sigma}_n)$$

4. Finally: get estimate of $\widehat{f}_Y$

$$\widehat{f}_Y(Y; \widehat{\Sigma}_n) = |\widehat{\Sigma}_n|^{-1/2} \widehat{g}_n\{(Y-\mu)^\top \Sigma^{-1}(Y-\mu); \widehat{\Sigma}_n\}$$

Covariance matrix estimation

- **Idea 1:** Factor estimator. Excess returns of a portfolio follow a factor model
- **Idea 2:** Shrinkage estimator. Estimator as a combination of biased and unbiased estimator (trade-off between a bias and an estimation error)

Factor Estimator

$$Y = B_n f + \varepsilon$$

- $Y = (Y_1, \dots, Y_p)^\top$ asset returns
- $B = (b_1, \dots, b_p)^\top$ factor loadings
  $b_i = (b_{n,i1}, \dots, b_{n,iK})$ $i = 1, \dots, p$
- $f = (f_1, \dots, f_K)^\top$ factors
- $\varepsilon = (\varepsilon_1, \dots, \varepsilon_p)^\top$ errors

OLS + diagonal covariance matrix of errors $\rightarrow \widehat{\Sigma}_{FFL}$

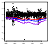
Shrinkage estimator

- Unbiased empirical variances $\sigma_{11}^2, \ldots, \sigma_{pp}^2$
- Shrinkage target: median value of all σ_i for diagonal elements (and 0 for other)
- Estimator:

$$\sigma_i^* = \widehat{\lambda}^* \sigma_{median} + (1 - \widehat{\lambda}^*)\sigma_i \tag{1}$$

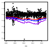
optimal pooling parameter $\widehat{\lambda}^*$

$$\widehat{\lambda}^* = \min\left(1, \frac{\sum_{k=1}^{p} \widehat{\mathrm{Var}(\sigma_k)}}{\sum_{k=1}^{p} (\sigma_k - \sigma_{median})^2}\right) \tag{2}$$

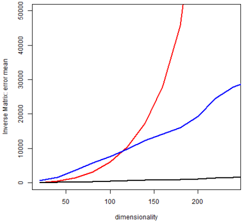


Simulation

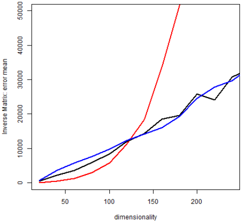
1. $K = 3$ factor model, generate normal sample of factors $n = 250$
2. p from 20 to 400 by 20
    2.1 Generate normal factor loading vectors
    2.2 Generate p standard deviations from Gamma distribution
    2.3 Generate normal error vectors
    2.4 Get sample of returns according to the model
3. Repeat M =1000 times



Inverse Matrix



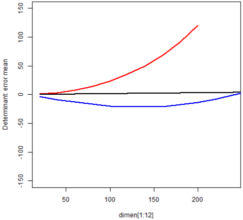
3 factors



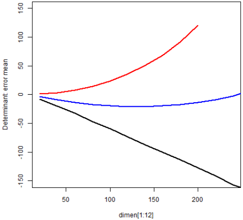
2 factors

Figure 1: The average error for $\widehat{\Sigma}_{shrink}$ (blue curve), $\widehat{\Sigma}_{FFL}$ (black curve) and $\widehat{\Sigma}_{sam}$ (red curve) under Frobenius norm plotted against dimensionality p, $n = 250$, $M = 1000$

Determinant



3 factors



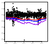
2 factors

Figure 2: Logarithm of the determinant of true covariance matrix divided by the determinant of the $\widehat{\Sigma}_{shrink}$ (solid blue curve), $\widehat{\Sigma}_{FFL}$ (black curve) and $\widehat{\Sigma}_{sam}$ (red curve) plotted against dimensionality p, $n = 250$, $M = 1000$ repetitions

Fama French 3 Factor Model and Carhart 4 Factor Model

$$Y_i = r_i - R_f = \alpha + \beta_1 (R_m - R_f) + \beta_2 SMB + \beta_3 HML + \beta_4 Mom \tag{3}$$

- R_f risk free rate (1-month TBill)
- R_m market rate (The NYSE Composite Index)
- SMB the performance of small stocks relative to big stocks (Small Minus Big)
- HML the performance of value stocks relative to growth stocks (High Minus Low)
- Mom momentum



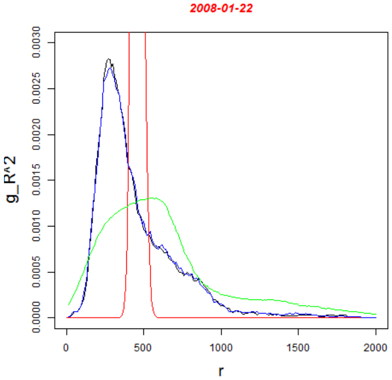


Figure 3: $g_{R^2}(r)$: for normal distribution and estimated with $\widehat{\Sigma}_{FFL}$ for **3 factors** and 4 factors and $\widehat{\Sigma}_{shrink}$ for S&P500 daily returns with monthly interval, $n = 750$, p=459

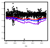
VaR

Profit and loss for a a linear portfolio $\Pi(t)$

$$\Delta\Pi(t) = \delta_1 X_1 + \ldots + \delta_p X_p(t)$$

VaR: $\mathrm{P}\{\Delta\Pi(t) < -VaR_\alpha\} = \alpha$ Assumptions:

- Returns $X = (X_1, \ldots, X_p)$ are elliptically distributed
- Weights $\delta = (\delta_1, \ldots, \delta_p)$ are known



VaR

Solve $\alpha = |\Sigma|^{-1/2} \int_{(\delta x \leq -VaR_\alpha)} g\{(x-\mu)^\top \Sigma^{-1} (x-\mu)\}\, dx$

$$VaR_\alpha = -\delta\mu + q^g_{\alpha,p} \sqrt{\delta^\top \Sigma \delta}$$

$s = q^g_{\alpha,p} \,:\, \alpha = G(s)$

$$G(s) = \frac{2\pi^{\frac{n-1}{2}}}{\Gamma(\frac{n-1}{2})} \int_s^\infty \int_{z_1^2}^\infty (u - z_1^2)^{\frac{n-3}{2}} g(u)\, du\, dz_1$$

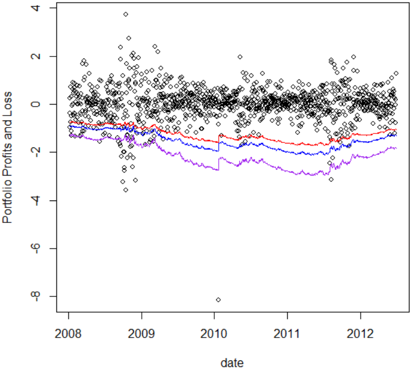


Figure 4: VaR estimated with $\widehat{\Sigma}_{FFL}$ for 3-factor model for the S&P500 portfolio for daily returns for 5% level, 2.5% level, 0.5% level $n = 750$, $p = 459$

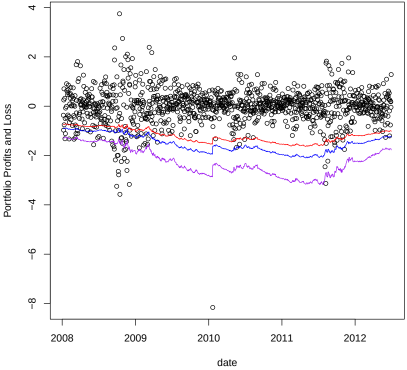


Figure 5: VaR estimated with $\widehat{\Sigma}_{FFL}$ for 4-factor model for the S&P500 portfolio for daily returns for 5% level, 2.5% level, 0.5% level $n = 750$, $p = 459$

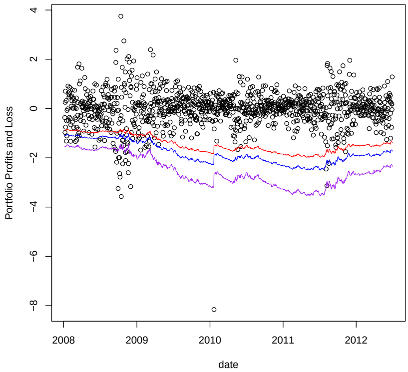


Figure 6: VaR estimated with $\widehat{\Sigma}_{shrink}$ for the S&P500 portfolio for daily returns for 5% level, 2.5% level, 0.5% level $n = 750$, $p = 459$

α	5%	2.5%	0.5%
3 *factors*	7.8%	5.1%	2.3%
4 *factors*	7.1%	4.8%	2.1%
shrinkage	5.3%	3.8%	1.7%

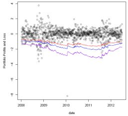
Table 1: Theoretical quantiles and percentage of outliers

α	5%	2.5%	0.5%
3 *factors*	3.0%	1.7%	0.4%
4 *factors*	3.2%	1.7%	0.2%
shrinkage	1.9%	1.0%	0.2%

Table 2: Theoretical quantiles and percentage of outliers excluding crisis

Elliptical Distributions in High Dimensions

Wolfgang Karl Härdle
Ostap Okhrin
Irina Pimenova



Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
http://lvb.wiwi.hu-berlin.de

References

J. Fan, W. Härdle and O. Okhrin
Semiparametric Estimation for very High Dimensional Elliptical Distributions
2012

J. Fan, Y. Fan, and J. Lv
High dimensional covariance matrix estimation using a factor model
Journal of Econometrics, 147, 186-197. 2008

E. Liebscher
A semiparametric density estimator based on elliptical distributions
Journal of multivariate analysis, 92, 205-225. 2005

References

J. Kamdem
Value-at-Risk and expected shortfall for linear portfolios with elliptically distributed risk factors
International Journal of Theoretical and Applied Finance, 8, 537-551. 2005

E. Fama and K. French
Common risk factors in the returns on stocks and bonds
Journal of Financial economics, 33, 3-56. 1993

References

R. Opgen-Rhein and K. Strimmer
ccurate ranking of differentially expressed genes by a distribution-free shrinkage approach
Stat Appl Genet Mol Biol, 6: Article 9. 2007

J. Bien and R. Tibshirani
Sparse estimation of a covariance matrix
Biometrika, 2010