Issues in large covariance estimation for portfolio risk prediction

Stjepan Begušić (stjepan.begusic@fer.hr)

Nov 29, 2022

Laboratory for Financial and Risk Analytics (lafra.fer.hr)
Faculty of Electrical Engineering and Computing
University of Zagreb, Croatia

Laboratory for Financial and Risk Analytics

Introduction

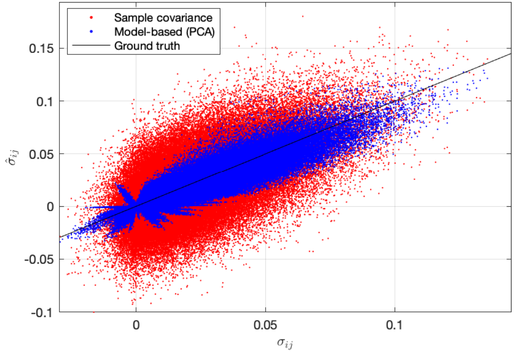
Covariance estimation and applications in finance

Portfolio optimization:

- Positive definiteness is crucial
- Mitigate the "error maximization" property: bias-variance tradeoff

Portfolio risk prediction:

- Positive definiteness not necessary
- Minimize estimation noise - get closer to the "ground truth"



What is considered in this experiment:

- A number of model-based estimators (POET, S-POET, JSE, Ledoit-Wolf)
- Varying dimensionality of the data sample (p and n)
- Different portfolios, not optimized using the covariance estimate (exogenous)
- Focus on portfolio variance prediction errors

Model

- Random vector of observable variables $\mathbf{X} = [X_1, X_2, ..., X_p]^\intercal$
- Unobservable common factor $F \in \mathbb{R}$, with variance σ^2
- Unobservable factor loadings $\beta = [\beta_1, \ldots, \beta_p]^\intercal$
- Unobservable specific factors $\varepsilon = [\varepsilon_1, \varepsilon_2, ..., \varepsilon_p]^\intercal$, with diagonal covariance Ψ

Model:

$$\mathbf{X} = \beta F + \varepsilon.$$

Covariance under the model:

$$\Sigma = \sigma^2 \beta \beta^T + \Psi,$$

with largest eigenvalue λ and corresponding eigenvector h.

Sample estimation and preliminaries

- Data sample $X \in \mathbb{R}^{n \times p}$
- Sample covariance $\hat{\Sigma}$ with largest eigenvalue $\hat{\lambda}$ and corresponding eigenvector $\hat{h}$
- Portfolio w with variance $w^\intercal \Sigma w$ estimated by $w^\intercal \hat{\Sigma} w$
- Depending on w, what estimators work best for Σ, under different p and n?

Estimators

Principal orthogonal complement thresholding (POET)[1] under a 1-factor model:

$$\hat{\Sigma}_{POET} = \hat{\lambda}\hat{h}\hat{h}^{\intercal} + \hat{\Psi},$$

where $\tilde{\Psi} = \hat{\Sigma} - \hat{\lambda}\hat{h}\hat{h}^{\intercal}$ and $\hat{\Psi} = (\hat{\psi}_{ij})_{p \times p}, \quad \hat{\psi}_{ij} = \begin{cases} \tilde{\psi}_{ii}, & i = j, \\ s_{\tau_{ij}}(\tilde{\psi}_{ij}), & i \neq j. \end{cases}$

Properties:

- The trace of $\hat{\Sigma}$ is preserved,
- Here a diagonal $\hat{\Psi}$ estimator is considered ($\tau_{ij} = \infty, \ \forall i, j$).

[1] Fan, J., Liao, Y., & Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 75(4), 603–680.

Shrinking the eigenvalues

- Sample eigenvalues are biased[2].
- S-POET – shrink the eigenvalue estimate:

$$\hat{\lambda}_S = \hat{\lambda} - \frac{p}{n} \frac{\mathrm{Tr}(\hat{\Sigma}) - \hat{\lambda}}{(p - 1 - p/n)}$$

- S-POET estimator:

$$\hat{\Sigma}_{SPOET} = \hat{\lambda}_S \hat{h} \hat{h}^\intercal + \hat{\Psi}$$

[2]Wang, W., & Fan, J. (2017). Asymptotics of empirical eigenstructure for high dimensional spiked covariance. Annals of Statistics, 45(3), 1342–1374.

Shrinking the eigenvector

- Eigenvector mean and dispersion:

$$\mu_h = 1/p \sum_{i=1}^{p} h_i, \quad d_h^2 = 1/p \sum_{i=1}^{p} (h_i/\mu_h - 1)^2$$

- Sample eigenvectors have higher dispersion and lower mean[3].

[3] Goldberg, L. R., Papanicolaou, A., & Shkolnik, A. (2022). The Dispersion Bias. SIAM Journal on Financial Mathematics, 13(2), 521–550.

JSE estimator

- JSE correction[4] - reduce the dispersion bias:

$$\hat{h}_{JSE} = \frac{\mu_{\hat{h}}\mathbf{1} + c_{JSE}(\hat{h} - \mu_{\hat{h}}\mathbf{1})}{|\mu_{\hat{h}}\mathbf{1} + c_{JSE}(\hat{h} - \mu_{\hat{h}}\mathbf{1})|}$$

with: $c_{JSE} = 1 - \frac{\nu^2}{s^2(\hat{h})}, \quad \nu^2 = \frac{\mathrm{tr}(\hat{\Sigma}) - \hat{\lambda}}{p(n-1)}, \quad s^2(h) = \frac{1}{p}\sum_{i=1}^{p}\left(\hat{\lambda}\hat{h}_i - \hat{\lambda}\mu_{\hat{h}}\right)^2$

- POET-JSE estimator:

$$\hat{\Sigma}_{POET-JSE} = \hat{\lambda}\hat{h}_{JSE}\hat{h}_{JSE}^{\intercal} + \hat{\Psi}$$

- SPOET-JSE estimator:

$$\hat{\Sigma}_{POET-JSE} = \hat{\lambda}_S\hat{h}_{JSE}\hat{h}_{JSE}^{\intercal} + \hat{\Psi}$$

[4]Goldberg, L. R., & Kercheval, A. (2022). James-Stein for the Leading Eigenvector.

Ledoit-Wolf shrinkage estimator

Shrinkage target[5]:

$$\hat{\Sigma}_T = \hat{\sigma}_T^2 \hat{\beta}_T \hat{\beta}_T^\intercal + \hat{\Psi}_T$$

with $\hat{F}_t = \frac{1}{p} \sum_i^p X_{ti}$, and the parameters $\hat{\sigma}_T$, $\hat{\beta}_T$ and $\hat{\Psi}_T$ estimated by OLS.

Ledoit-Wolf estimate:

$$\hat{\Sigma}_{LW} = \gamma \hat{\Sigma}_T + (1 - \gamma) \hat{\Sigma}$$

with a data-driven shrinkage intensity γ.

[5] Ledoit, O., & Wolf, M. (2022). The Power of (Non-)Linear Shrinking: A Review and Guide to Covariance Matrix Estimation. Journal of Financial Econometrics, 20(1), 187–218.

Experimental setup

Experimental setup

Data generating process: $\mathbf{X} = \beta F + \varepsilon$.

Factor exposures β_i	$\mathcal{N}(1, 0.25)$
Factor returns F	$\mathcal{N}(0, 0.16^2)$
Idiosyncratic returns ε_i	$T(5)$, mean 0 and variance 0.60^2
Number of variables p	$[100, 200, ..., 1000]$
Sample size n	$[100, 200, ..., 1000]$
Number of simulations N_s	100

For each simulation:

- Generate the model (factor exposures),
- Simulate the data matrix X with maximum p and n,
- Slice the data to obtain the sample for all p and n.

Performance measures

Frobenius norm:

$$FN_{\hat{\Sigma}} = ||(\Sigma - \hat{\Sigma})||_F$$

Portfolio risk prediction errors:

- Variance forecast ratio:

$$VFR_{w,\hat{\Sigma}} = \frac{w^\intercal \hat{\Sigma} w}{w^\intercal \Sigma w}$$

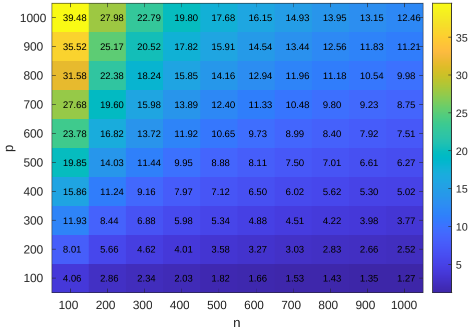
- Relative error:

$$RE_{w,\hat{\Sigma}} = |VFR_{w,\hat{\Sigma}} - 1|$$

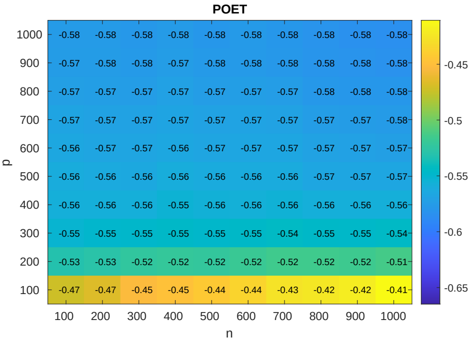
Results

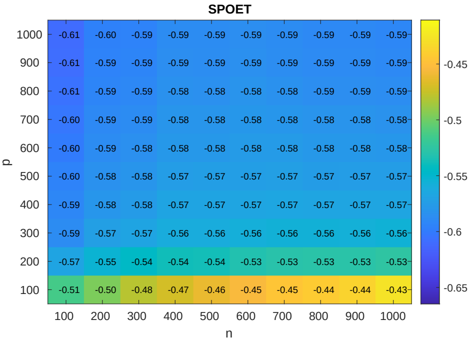
Frobenius norm of the error

Sample estimator, Frobenius norm of the error (w.r.t. the population covariance):

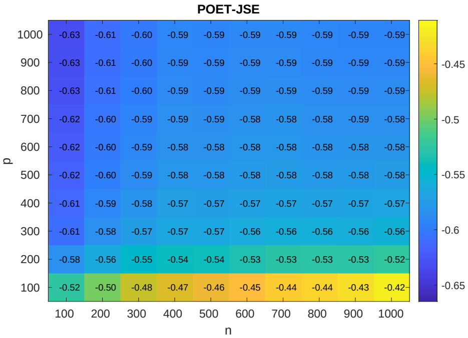


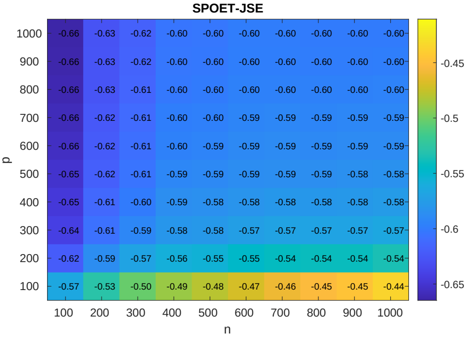
Reduction in error vs. sample estimator: $\frac{FN - FN_{sample}}{FN_{sample}}$





Frobenius norm of the error





- The model-based estimators seem to drastically reduce estimation error (as measured by the Frobenius norm of the error).

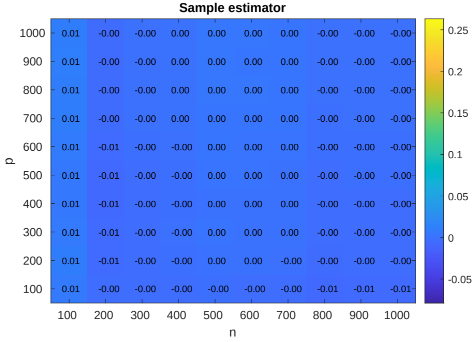
Equal weighted portfolio

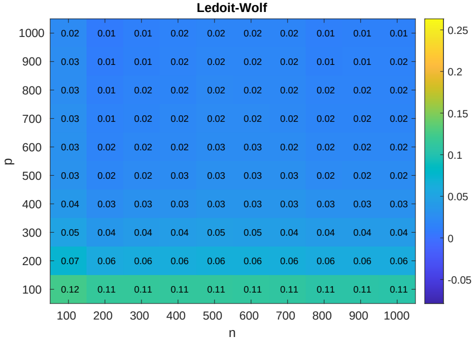
Equal weighted portfolio

Equal-weighted portfolio:

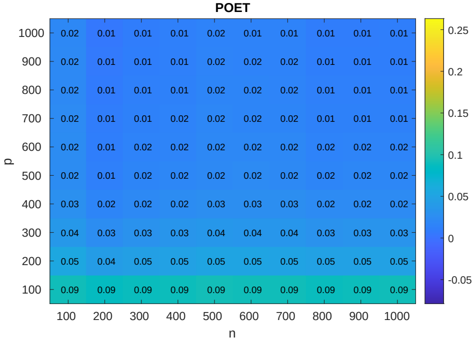
$$w_{EW} = [1/p, 1/p, ..., 1/p]^\intercal$$

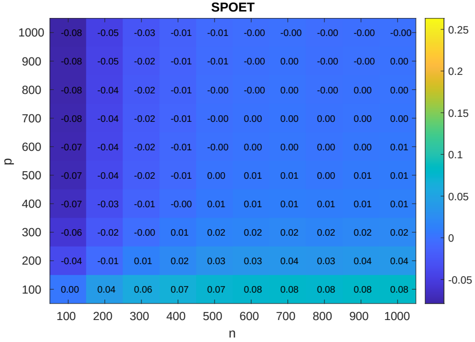
Average $VFR - 1$, across all simulations:





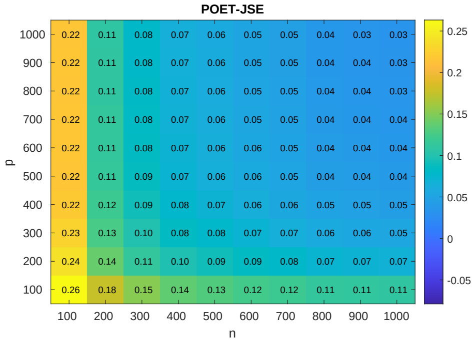
Equal weighted portfolio - variance forecast ratio

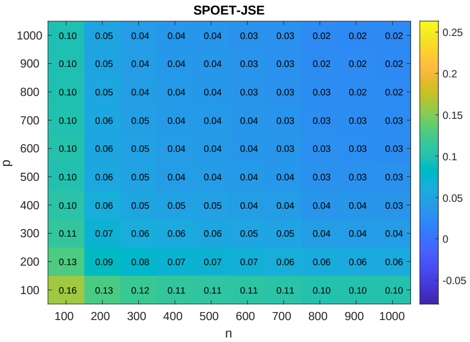




- Correcting the eigenvalue introduces negative bias into risk predictions?

Equal weighted portfolio - variance forecast ratio





- Correcting the eigenvector introduces positive bias into risk predictions?

Portfolio risk decomposition

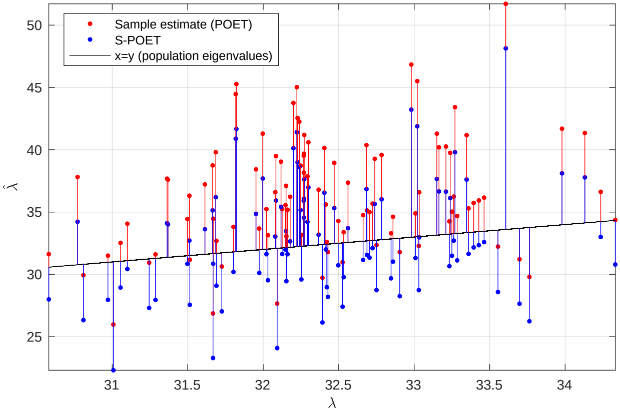
Portfolio variance under the model:

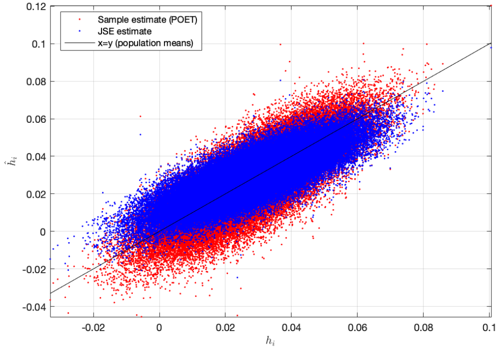
$$\sigma_w^2 = w^T \Sigma w = \sigma^2 (w^T \beta)^2 + w^T \Psi w = \sigma^2 |\beta|^2 (w^T h)^2 + w^T \Psi w$$

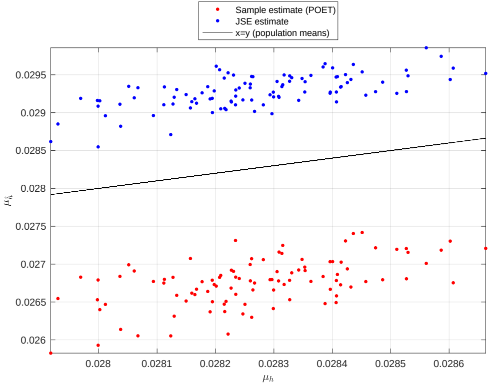
Equal weighted portfolio is a special case: $(w^T h)^2 = \mu_h^2$

- The only important quantities in the systematic component are the eigenvalue and eigenvector mean (but not its direction)!

Eigenvalue correction







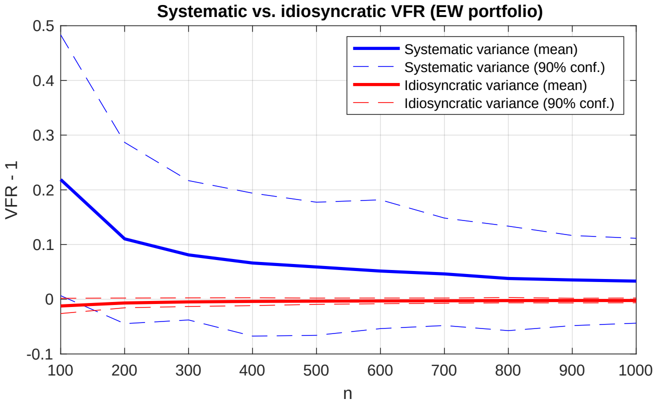
Is there a ”better” shrinkage intensity estimate?

- Cross-validation or bootstrapping do not seem to provide an answer.

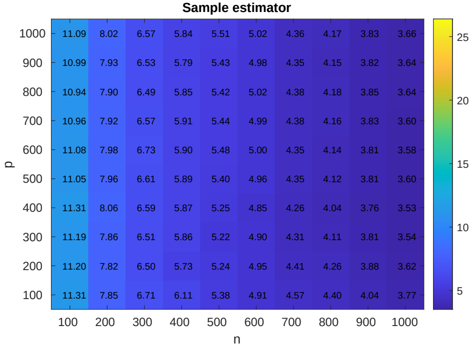
Systematic and idiosyncratic variance errors - POET-JSE

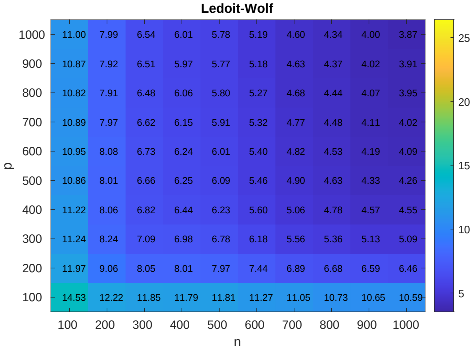
Systematic and idiosyncratic variance VFR:

$$VFR_{SV} = \frac{w^\intercal(\hat{\lambda}\hat{b}\hat{b}^\intercal)w}{w^\intercal(\lambda b b^\intercal)w}, \quad VFR_{IV} = \frac{w^\intercal\hat{\Psi}w}{w^\intercal\Psi w}$$

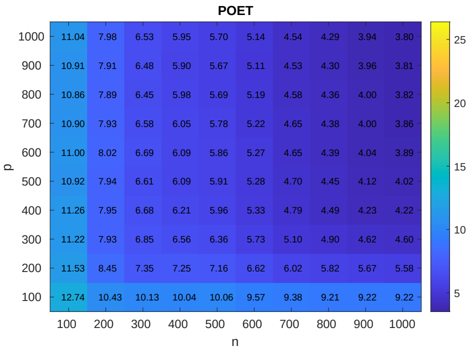


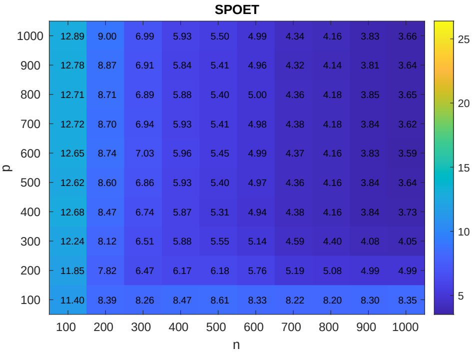
Average RE (in %), across all simulations:



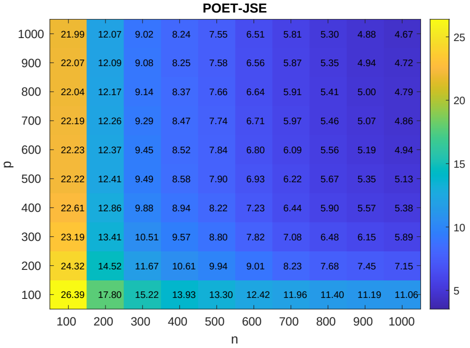


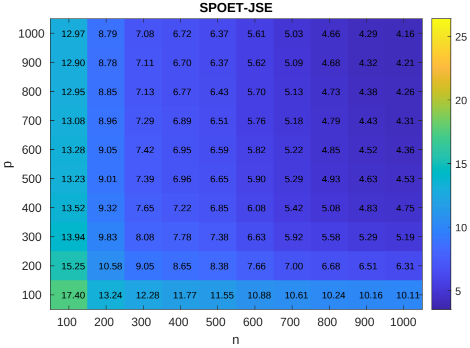
Equal weighted portfolio - relative error





Equal weighted portfolio - relative error





- The RE of the sample estimator depends mostly on the sample size n, rather than p
- There seems to be no gain from model-based estimators in measuring the variance of the EW portfolio.

Capitalization weighted portfolio

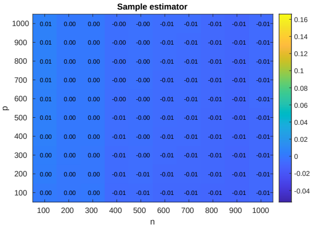
Marketcap portfolio

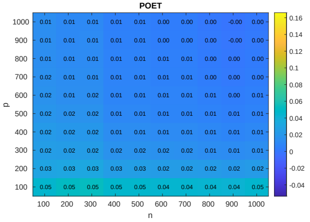
Capitalization-weighted portfolio:

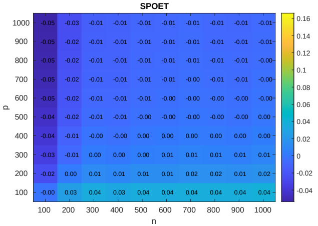
$$w_{MC} = \frac{\tilde{w}}{\sum \tilde{w}}, \quad \tilde{w}_i = \frac{1}{i}, \quad i = 1, ..., p$$

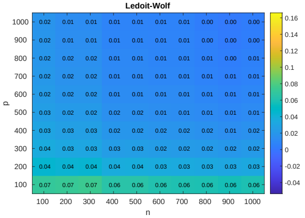
Marketcap portfolio - variance forecast ratio

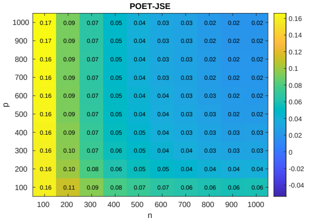
Average $VFR - 1$, across all simulations:

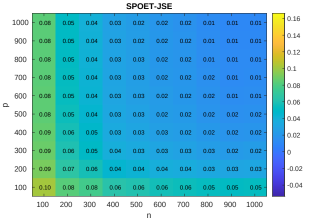






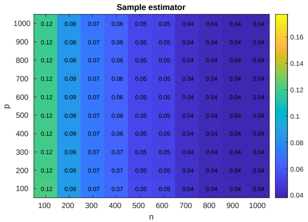


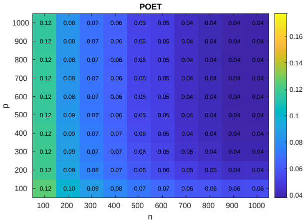


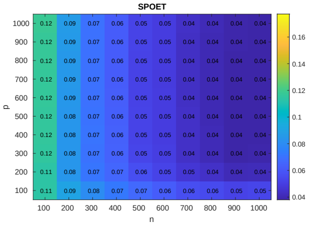


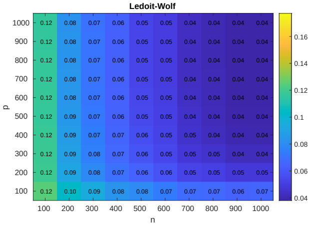
Marketcap portfolio - relative error

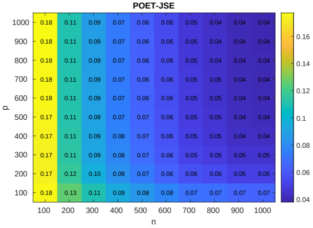
Average RE (in %), across all simulations:

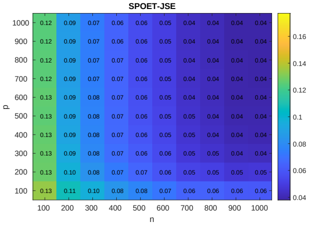












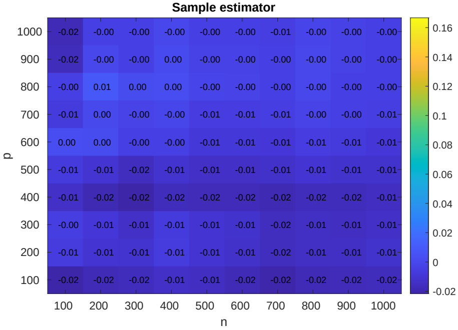
Minimum variance portfolio

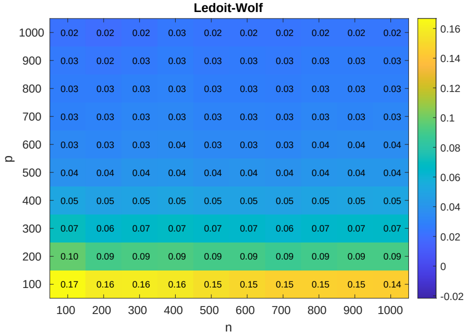
Minimum variance portfolio

Minimum variance portfolio (using the population covariance):

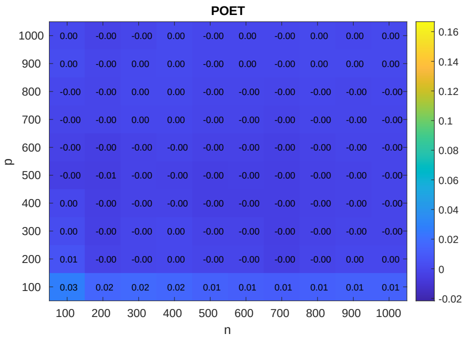
$$w_{GMV} = \frac{\Sigma^{-1}\mathbf{1}}{\mathbf{1}^\top\Sigma^{-1}\mathbf{1}}$$

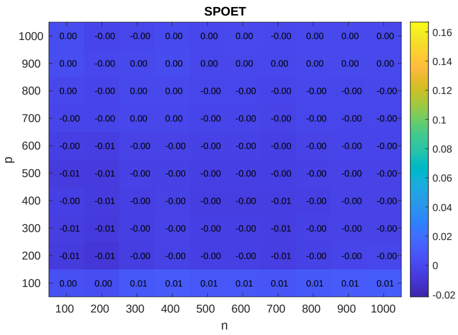
Average $VFR - 1$, across all simulations:



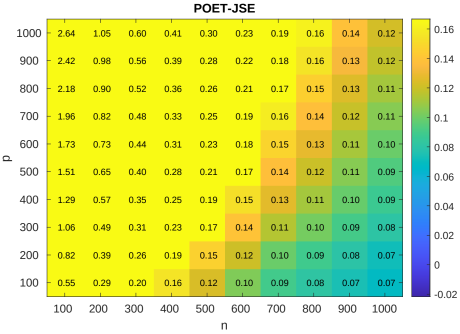


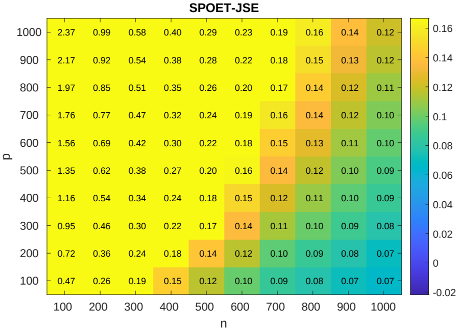
Minimum variance portfolio - variance forecast ratio



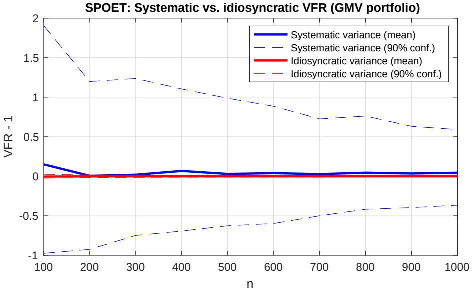


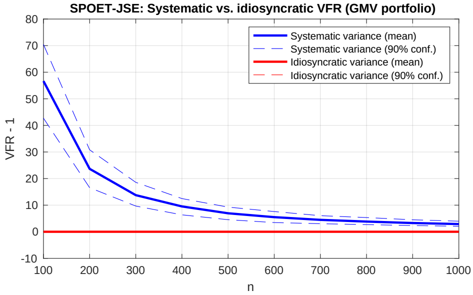
Minimum variance portfolio - variance forecast ratio





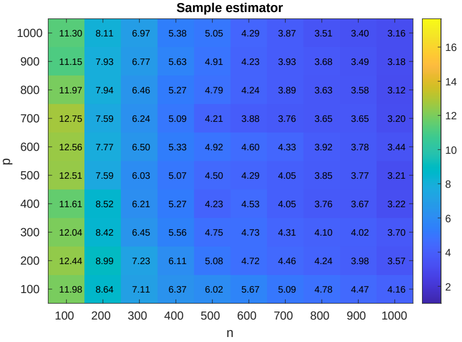
- In the simulated scenario, the minimum variance portfolio has $\sim 5\%$ systematic and $\sim 95\%$ idiosyncratic risk (for $p = 1000$).
- The variance forecast ratios for these components are:

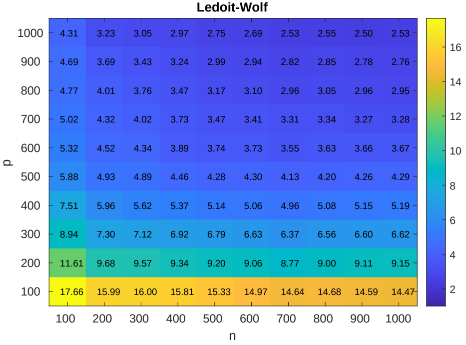




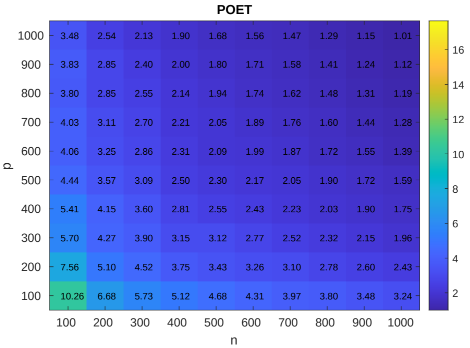
Minimum variance portfolio - relative error

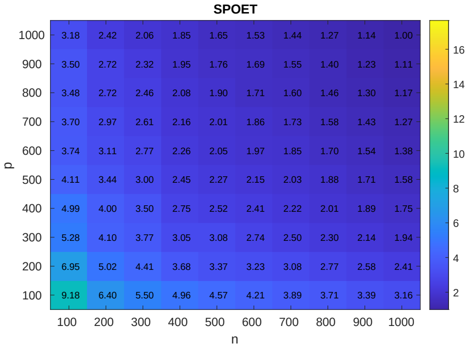
Average RE (in %), across all simulations:



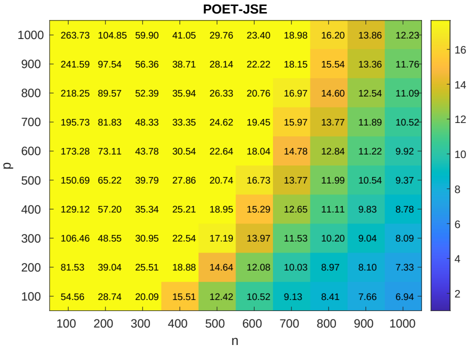


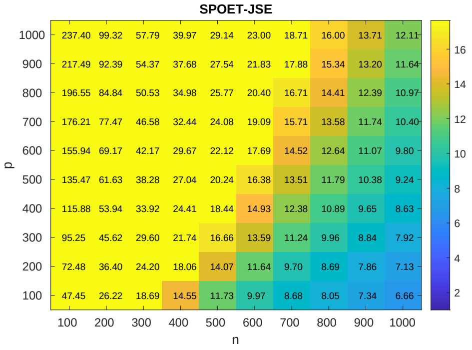
Minimum variance portfolio - relative error





Minimum variance portfolio - relative error





Concluding remarks

Concluding remarks

- Eigenvalue shrinkage helps with the bias, but not with the estimation noise.
- Eigenvector shrinkage eliminates optimization bias, but introduces eigenvector bias.
- Sample estimation errors mostly depend on data sample sizes - sample covariance remains a difficult benchmark to beat!
- In HDLSS the estimation error can be greatly reduced for some portfolios.