James-Stein Empirics

WRDS US equity data set study.

Jacob Bien & Alex Shkolnik

JOINT FALL SEMINAR

October 10, 2024.

Department of Statistics & Applied Probability
University of California, Santa Barbara

The data set

Wharton Research Data Services (WRDS).

https://wrds-www.wharton.upenn.edu

- *Access through UCLA library.*
- *2003-2023 time-series of US equity returns (+ market caps).*
- *The frequency is daily (return).*
- *There is missing data (e.g., acquisition, merger, bankruptcy).*
- *We study the 3000 stocks with the largest market cap.*
- *The constituents of this group changes over time.*

The testing framework

We observe a vector $r_j \in \mathbb{R}^p$ on date j.

- *p is the number of stocks/securities/assets.*
- $r_j = (r_{1j}, \ldots, r_{pj})^\top$
- *We observe r_j on n dates.*
- *$(p \times n)$ data matrix $R = (r_{ij})_{1 \le i \le p, 1 \le j \le n}$.*

r_{ij} is the return of stock i on date j.

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pn} \end{pmatrix}.$$

Observing R we construct a portfolio $w \in \mathbb{R}^p$.

$$w = (w_1, \ldots, w_p)$$

- *w_i is the investment is stock i.*
- $\sum_{i=1}^{p} w_i = 1$ *(w.l.o.g)*
- $w_i \geq 0$ *(long position) and* $w_i < 0$ *(short position).*

Observing R we construct a portfolio $w \in \mathbb{R}^p$.

$$w = (w_1, \ldots, w_p)$$

- *$w_i \in \mathbb{R}$ is the investment is stock i with $\sum_{i=1}^{p} w_i = 1$.*

In-sample portfolio return

- *For the n column of R, i.e., r_n, we can compute*

$$r_{w\text{-}in} = \langle r_n, w \rangle = \textstyle\sum_{i=1}^{p} r_{in} w_i \, .$$

- *This is the* **in-sample portfolio return** *and we are in full control of this number (i.e. all of R is available).*

Out-of-sample portfolio return

- *Fixing w, we wait for some period m and compute the return,*

$$r_{w\text{-}out} = \langle r_{n+m}, w \rangle = \textstyle\sum_{i=1}^{p} r_{i(n+m)} w_i \, .$$

- *This is the* **out-of-sample portfolio return** *and we have no control of this number (i.e., r_{n+m} is not observed at time n).*

Both $r_{w\text{-in}}$ and $r_{w\text{-out}}$ are random variables (RVs).

- $r_{w\text{-}out}$ *may be viewed as a RV conditional on* $\boldsymbol{R}$.
- $r_{w\text{-}in}$ *may be viewed as a RV that is a function of* $\boldsymbol{R}$.

Both a mean, variance, . . . , distribution (histogram).

Using historical data, we can obtain a time-series for each.

- $r^{(1)}_{w\text{-}out}, r^{(2)}_{w\text{-}out}, r^{(3)}_{w\text{-}out}, \ldots, r^{(N)}_{w\text{-}out}$ *and same for* $r_{w\text{-}in}$.
- *This will depends on the choices of* p, n *and* m.
- p *is the number of variables (stocks).*
- n *is the observation window (training data) size.*
- m *is step size the window is shifted by.*
- *These matter! e.g.,* $m \geq n$, *the windows are not overlapping.*
- $p > n$ *vs* $p \leq n$.
- *size of* n *is related to "stationarity" assumptions on the data.*

The first window that leads to $r_{w\text{-out}}^{(1)}$ and $r_{w\text{-in}}^{(1)}$.

$$R^{(1)} = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pn} \end{pmatrix}.$$

The second window (shifted by m) that leads to $r_{w\text{-out}}^{(2)}$ and $r_{w\text{-in}}^{(2)}$.

$$R^{(2)} = \begin{pmatrix} r_{1(m+1)} & r_{1(m+2)} & \cdots & r_{1(n+m)} \\ r_{2(m+1)} & r_{2(m+2)} & \cdots & r_{2(n+m)} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ r_{p(m+1)} & r_{p(m+2)} & \cdots & r_{p(n+m)} \end{pmatrix}.$$

Testing framework inputs.

- *Historical data (e.g., WRDS).*
- p, n, m *and* M *(method to compute portfolio weights* w*).*
- *An example of* M *is principal-component analysis and mean-variance optimization (both upcoming).*

Testing framework outputs.

- $r^{(1)}_{w\text{-}out}, r^{(2)}_{w\text{-}out}, r^{(3)}_{w\text{-}out}, \ldots, r^{(N)}_{w\text{-}out}$ *and same for* $r_{w\text{-}in}$.
- *Some metric that evaluates the performance of* M.
- *Example metrics are the out-of-sample return and variance.*

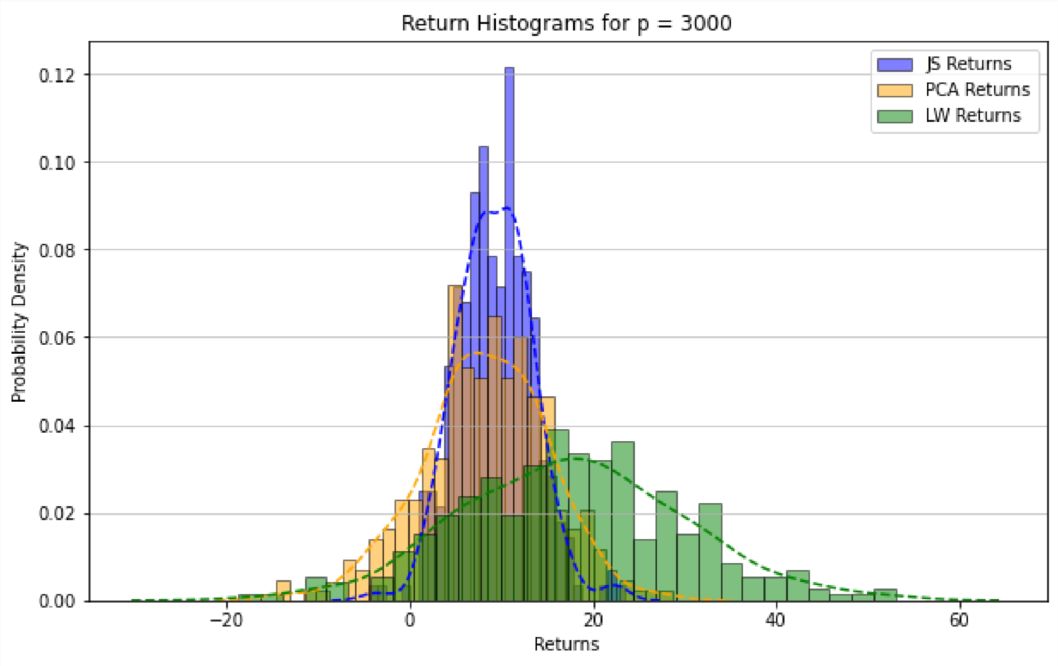
$$\mu_{w\text{-}out} = \frac{1}{N}\sum_{\ell=1}^{N} r^{(\ell)}_{w\text{-}out}, \qquad \sigma^2_{w\text{-}out} = \frac{1}{N-1}\sum_{\ell=1}^{N}(r^{(\ell)}_{w\text{-}out} - \mu_{w\text{-}out})^2$$

and their “running” versions as time-series.

Benchmarks portfolios.

- *Equally weighted portfolio, i.e.* $w_i = 1/p$.
- *Market cap weighted portfolio, i.e.,* $w_i = cap_i / \sum_{i=1}^{p} cap_i$

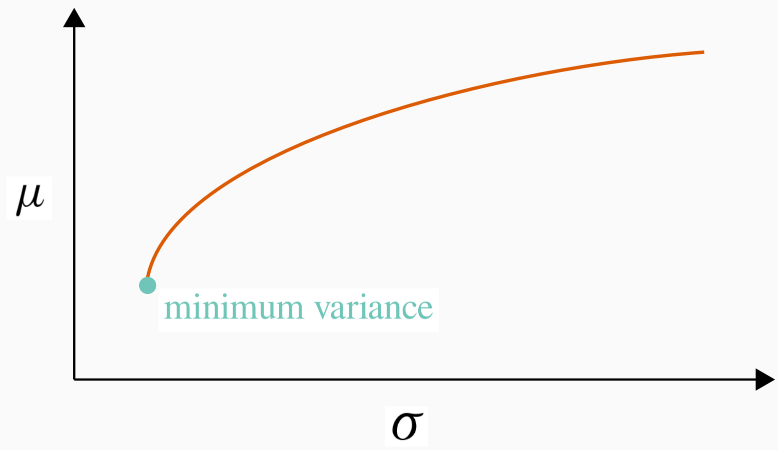
Histogram of $r_{w\text{-out}}^{(1)}, r_{w\text{-out}}^{(2)}, r_{w\text{-out}}^{(3)}, \ldots, r_{w\text{-out}}^{(N)}$ for 3 methods M.



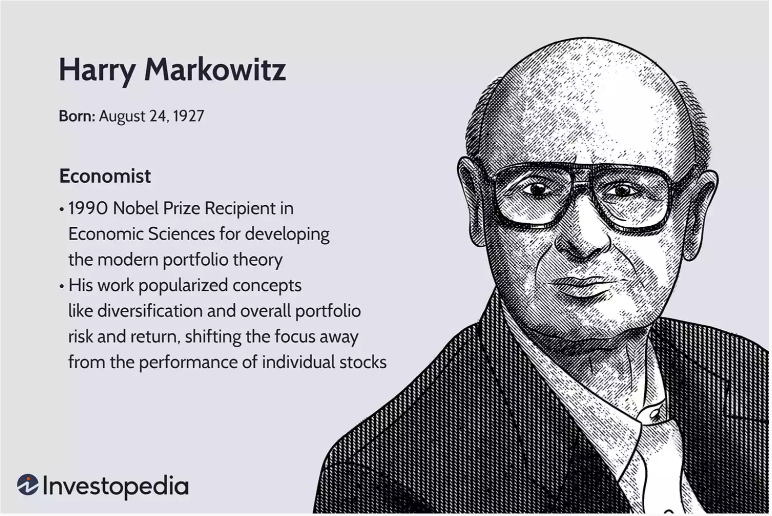
This is simulated (not empirical) data!

Mean-variance optimization

Since Markowitz (1952), quantitative investors have constructed portfolios with mean-variance optimization.



- *A simple quadratic program given a covariance matrix Σ.*
- *We can make two curves (in-sample and out-of-sample).*



Portfolio Selection Revisited (2024)

In honor of Harry Markowitz, 1927–2023.

The Markowitz quadratic program.

$$\min_{w \in \mathbb{R}^p} \langle w, \Sigma w \rangle$$

subject to:

$$\langle m, w \rangle \geq \alpha\,,$$

$$\langle \mathrm{e}, w \rangle = 1\,.$$

$$\text{(every } w_i \geq 0$$

$$\ldots \text{ etc.)}$$

- $\langle x, y \rangle = \sum_{i=1}^{p} x_i y_i$.
- Σ *is a* $(p \times p)$ *covariance matrix of stock returns.*
- $m \in \mathbb{R}^p$ *is the estimate of expected returns.*
- $\alpha \in \mathbb{R}$ *is the target portfolio return.*
- $\mathrm{e} = (1, \ldots, 1) \in \mathbb{R}^p$

The Markowitz quadratic program.

$$\min_{w \in \mathbb{R}^p} \langle w, \Sigma w \rangle$$

$$\text{subject to:}$$

$$\langle m, w \rangle \geq \alpha\,,$$

$$\langle \mathrm{e}, w \rangle = 1\,.$$

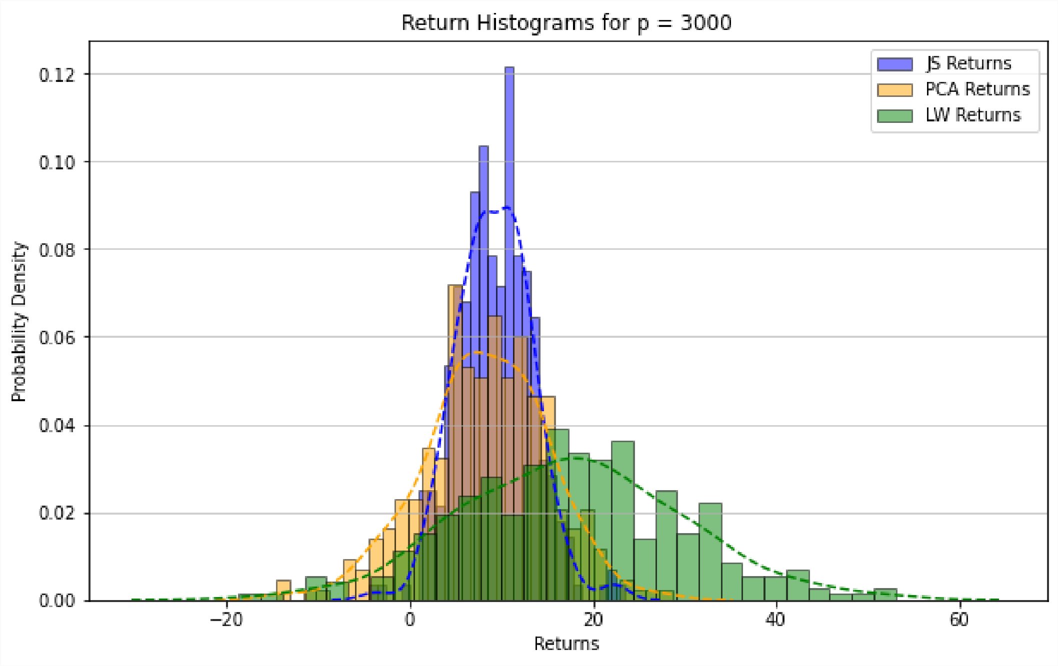
$$(\text{every } w_i \geq 0$$

$$\ldots \text{ etc.})$$

The Markowitz optimization enigma entails the observation that “*mean-variance optimizers are, in a fundamental sense, estimation-error maximizers*” – Michaud (1989).

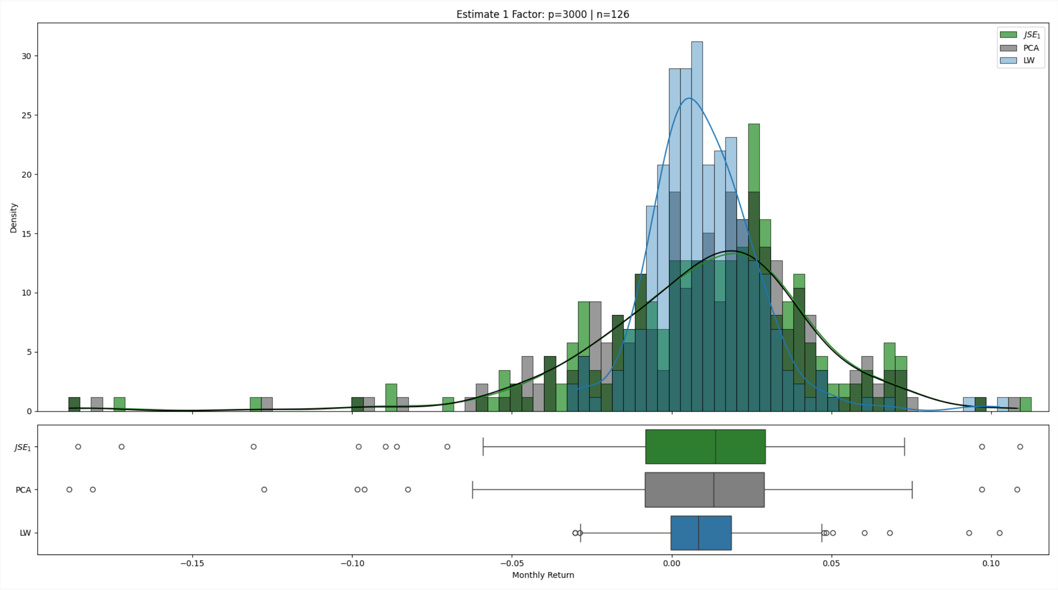
- *The estimation error sits in m and Σ.*

Histogram of $r_{w\text{-out}}^{(1)}, r_{w\text{-out}}^{(2)}, r_{w\text{-out}}^{(3)}, \ldots, r_{w\text{-out}}^{(N)}$ for 3 methods M.



The target return here is $\alpha = 8.5$. The standard deviation of the optimal portfolio return (with knowledge of true Σ) is 2.78.
The histograms above have 4.11, 6.85 and 12.2.

Histogram of $r^{(1)}_{w\text{-out}}, r^{(2)}_{w\text{-out}}, r^{(3)}_{w\text{-out}}, \ldots, r^{(N)}_{w\text{-out}}$ for 3 methods M.



Similar methods run on empirical data! ($\alpha = -\infty$)

Simulated vs empirical data

In simulation, R follows some statistical model.

- *Factor model.*

$$R = \mu + B\mathcal{F}^\top + \mathcal{E}$$

  $\mu \in \mathbb{R}^p$ – *expected return.*

  $\mathcal{F} \in \mathbb{R}^{n \times K}$ – *factor returns.*

  $B \in \mathbb{R}^{p \times K}$ – *factor exposures.*

  $\mathcal{E} \in \mathbb{R}^{p \times n}$ – *idiosyncratic return (error).*

  In this model B *and* μ *are estimated from observations of* R.

  Distributional assumptions are needed for $\mathcal{F}$ *and* $\mathcal{E}$.

- *A graphical model is another example.*

Given $(p \times n)$ data matrix R.

- *The $(p \times p)$ sample covariance is $Q = RR^\top / n$ (uncentered).*
- *The sample mean is $\bar{r} \in \mathbb{R}^p$ (average of columns of R).*
- *Subtract $\bar{r}$ from each columns of R to obtain Y (a $(p \times n)$ centered data matrix).*
- *The $(p \times p)$ centered sample covariance is $S = YY^\top / n$.*

The statistical properties of S (or Q) are examined via the spectral decomposition (eigenvalues $\mathscr{s}^2$ and eigenvectors h).

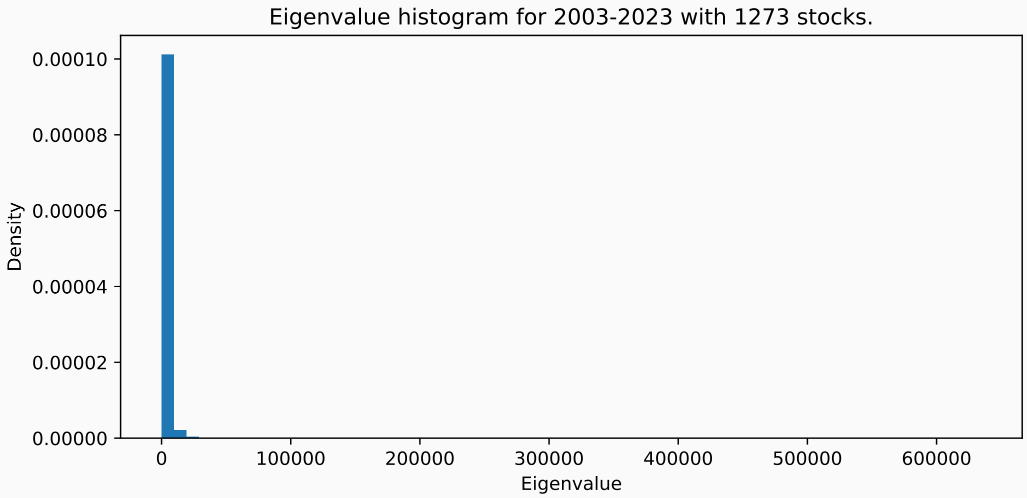
$$S = \sum_{(\mathscr{s}^2, h)} \mathscr{s}^2 h h^\top$$

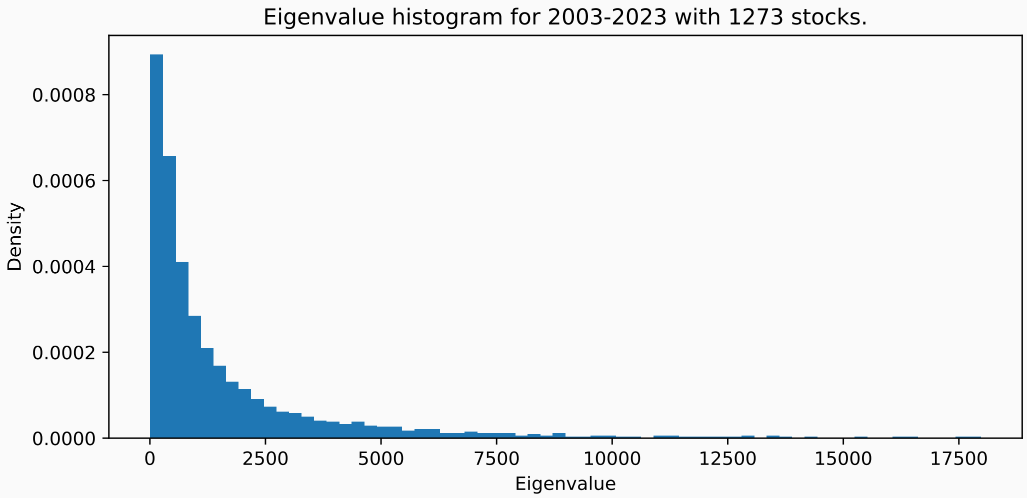
where $Sh = \mathscr{s}^2 h$ and h has unit length, $1 = |h|^2 = \langle h, h \rangle$.

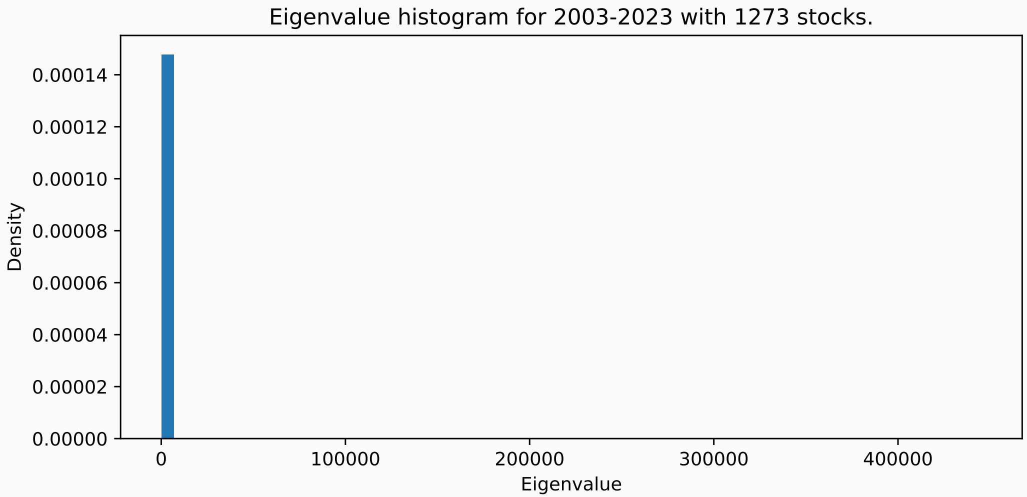
For Y (or R), the similar procedure may compute the singular value decomposition, e.g., $Y = \sum_{(\sigma, u, v)} \sigma u v^\top$.

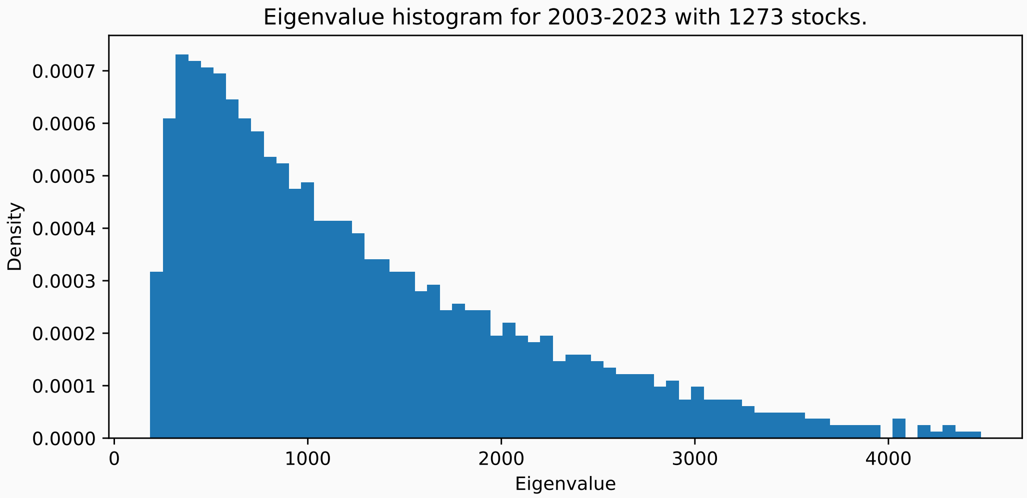
Useful empirics.

- *Compute all the eigenvalues of a simulated and empirical* S *to show the differences.*
- *Compute the eigenvector for the largest eigenvalue of a simulated and empirical* S *and plot the histogram of those entries.*
- *etc.*


Eigenvalue histogram for 2003-2023 with 1273 stocks.
Density
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
100000
200000
300000
400000
500000
600000
Eigenvalue


Eigenvalue histogram for 2003-2023 with 1273 stocks.
Density
0.0000
0.0002
0.0004
0.0006
0.0008
0
2500
5000
7500
10000
12500
15000
17500
Eigenvalue


Eigenvalue histogram for 2003-2023 with 1273 stocks.
Density
0.00014
0.00012
0.00010
0.00008
0.00006
0.00004
0.00002
0.00000
0
100000
200000
300000
400000
Eigenvalue


Eigenvalue histogram for 2003-2023 with 1273 stocks.
Density
0.0000
0.0001
0.0002
0.0003
0.0004
0.0005
0.0006
0.0007
0
1000
2000
3000
4000
Eigenvalue

Methods

Principal component analysis (PCA).

- *Starting with the spectral decomposition of S,*

$$S = \sum_{(\mathscr{s}^2, h)} \mathscr{s}^2 h h^\top = H H^\top + G$$

  where H is a $p \times K$ matrix (K principal components).

- *G is a $p \times p$ residual which is often regarded as noise on theoretical grounds (i.e. its eigenvalues are hypothesized to follow the Marchenko-Pastur distribution).*
- *Let $\Delta = \mathit{diag}(G)$, the matrix G with zeros off-diagonal.*
- *The PCA estimate of the covariance is given by*

$$\Sigma = H H^\top + \Delta.$$

  Note, H has K columns of the form $\eta = \mathscr{s} h$ where $\mathscr{s}^2$ is an eigenvalue of S with eigenvector h.

James-Stein for PCA (plain version).

– *We update the PCA estimate H as follows.*

$$H_{\text{JS}} = HC + F(I - C)$$

where $F = AA^{+}H$ for A^{+} the pseudo-inverse of A, any $p \times k$ matrix (for mean-variance we put the constraint vectors m and e as columns), and

$$C = I - \nu^2 J^{-1}, \quad J = (H - M)^{\top}(H - M),$$

for $\nu^2 = \frac{\mathrm{trace}(G)}{n_{+} - K}$ with n_{+} is the number of nonzero eigenvalues of the sample covariance S.

More refined versions to be added ...

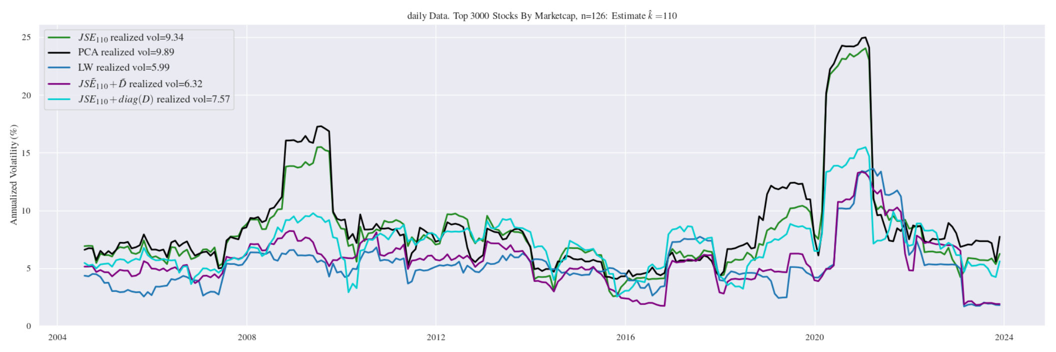
The Ledoit-Wolf family of estimators.

- *Starting with the sample covariance S we return*

$$\Sigma = c\,S + (1-c)\,F$$

  where F is a target matrix and c is a shrinkage intensity.
- $F = I$ *adjusts only the eigenvalues of* S.
- F, *constant correlation adjusts eigenvalues and eigenvectors (does well on empirical data but poorly in simulation).*

Running volatility plot for several methods.



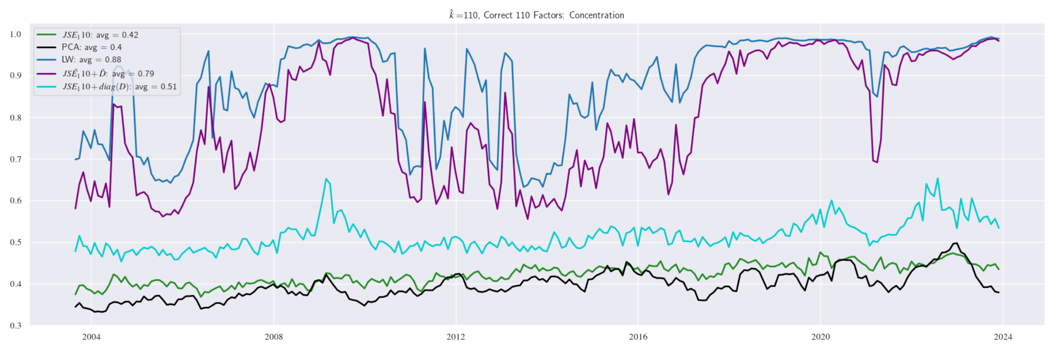
Theoretical assumptions vs practice

Metrics

Given $r_{w\text{-out}}^{(1)}, r_{w\text{-out}}^{(2)}, r_{w\text{-out}}^{(3)}, \ldots, r_{w\text{-out}}^{(N)}$ and same for $r_{w\text{-in}}$ we can look at the following metrics in- and out-of-sample.

- *Running mean and volatility.*
- *Sharpe and Sortino ratios.*
- *Concentration metrics (e.g., Herfindahl index).*
- *Portfolio turnover.*

Running concentration plot for several methods.



Pitfalls with empirical data

Besides code bugs that always come up . . .

- *Missing data (selection bias).*
- *Stock delisting and volatility (e.g., bankruptcy vs merger).*
- *Limited history (*1923– *with* 250 *trading days per year).*
- *Statistical properties of data is difficult to model (relevant for theoretical assumptions on methods + simulations).*

Related Literature

General approaches to mean-variance weights.

- *Covariance estimation.*
- *Robust optimization.*
- *Prorfolio weight shrinkage.*

Examples of empirical work.

- *Georgantas et al. (2024)*.

References

Georgantas, A., Doumpos, M. & Zopounidis, C. (2024), 'Robust optimization approaches for portfolio selection: a comparative analysis', *Annals of Operations Research* **339**(3), 1205–1221.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* **7**(1), 77–91.

Michaud, R. O. (1989), 'The markowitz optimization enigma: Is 'optimized'optimal?', *Financial analysts journal* **45**(1), 31–42.