Improved Covariance Matrix Estimation for Portfolio Risk Management

Henri Staal | Peregrine Capital | 26 March 2026

ACTUARIAL SOCIETY OF SOUTH AFRICA

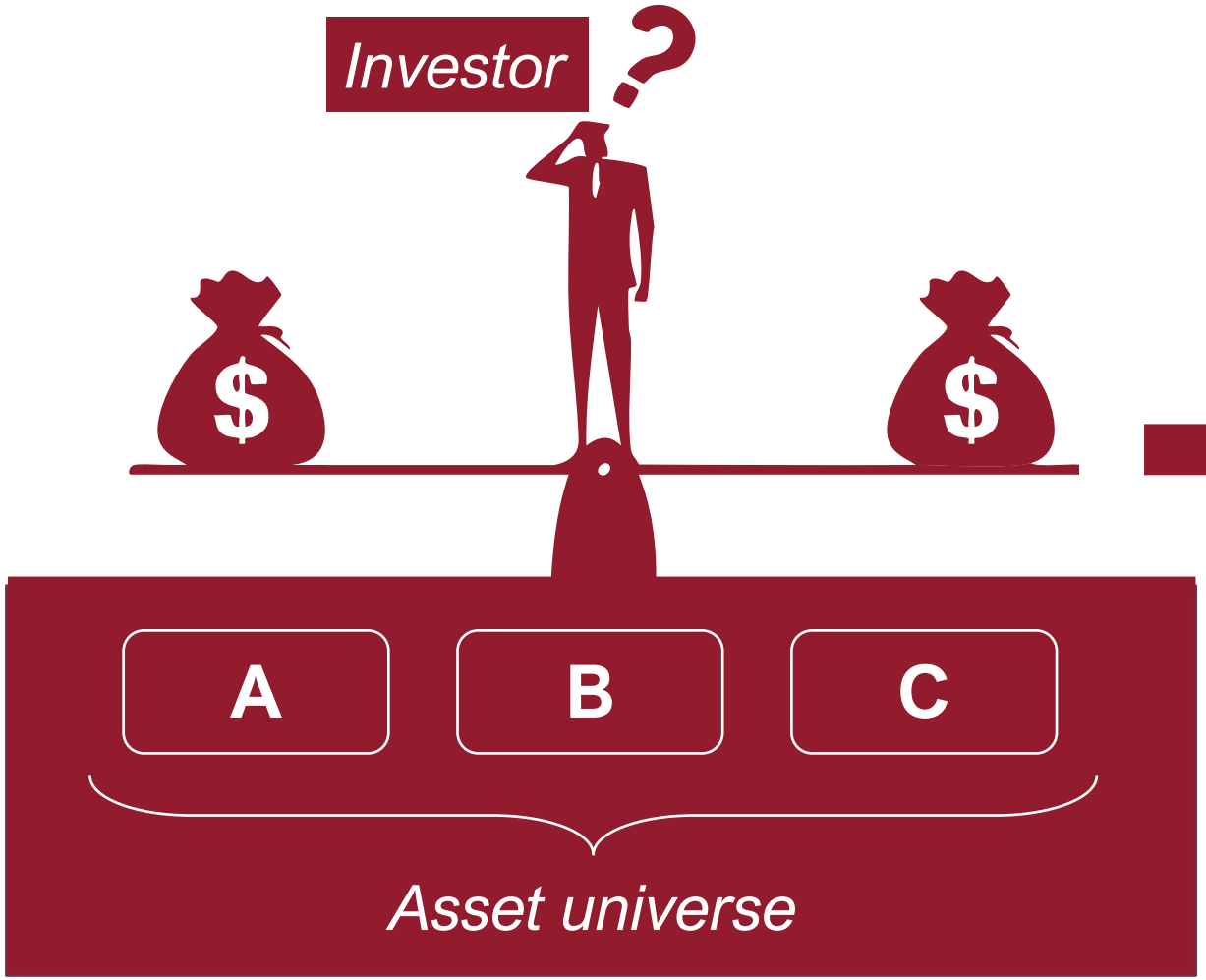
Advancing Actuarial Excellence

Educate • Develop • Uphold

Portfolio asset allocation

"Investors talk incessantly about what investment they think is a good buy. But rarely do investors openly discuss how much of that investment should be bought. This second question is of vital importance, yet very little literature exists to explain how individual ideas should be appropriately combined to deliver an effectively diversified portfolio." – Ensemble Capital

Portfolio asset allocation



Which asset allocation for **optimal balance** of **risk and return?**

Introduction and motivation- MVO

Markowitz (1952):

Mathematically Markowitz (1952) expressed as:

$\min(\mathbf{w}) \; \frac{1}{2}\mathbf{w}'\Sigma\mathbf{w}$

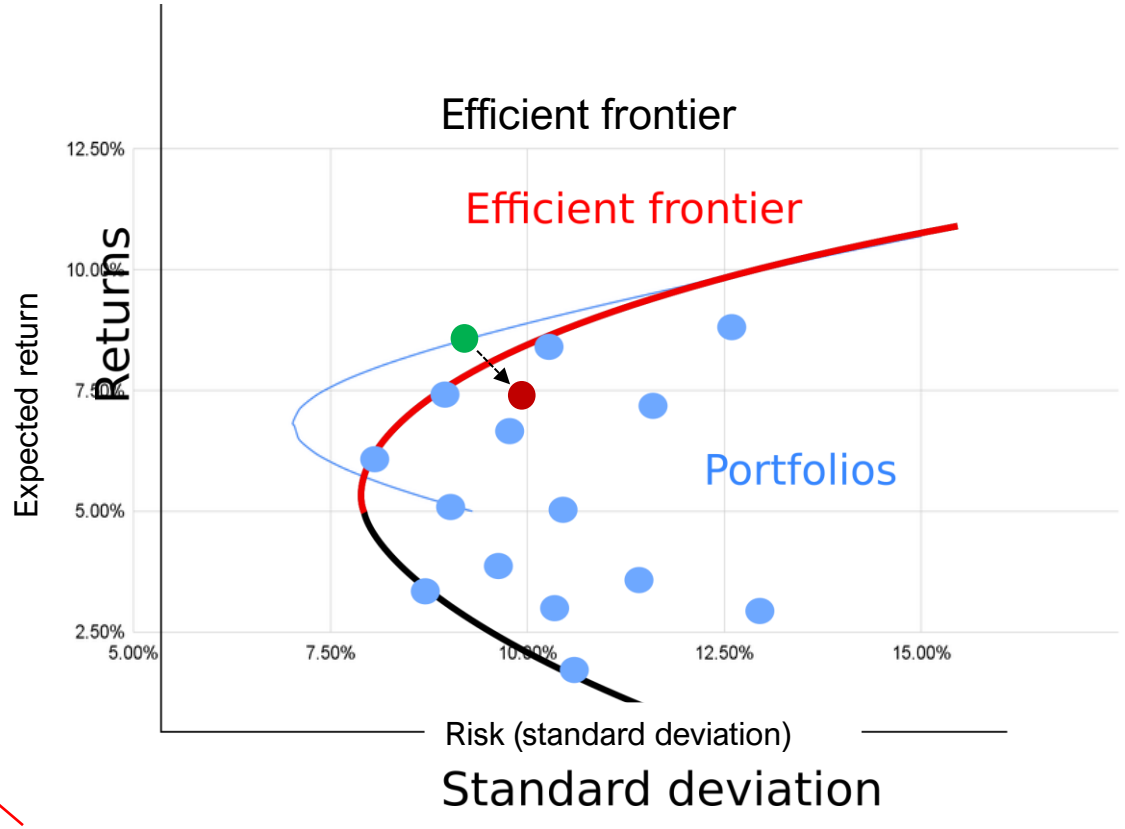
subject to $\mathbf{w}'\boldsymbol{\mu} = r$ and $\mathbf{w}'\mathbf{1} = 1$

Mathematical framework:

- **w** = weight vector
- **Σ** = covariance matrix
- **μ** = expected returns

Where:

- ***minimise volatility***
- ***given a target return***



How do we calculate?

Markowitz (1952) framework:

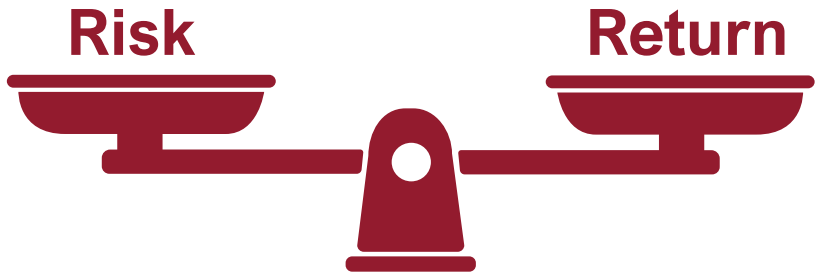
- Review of contemporary methods
- Developed a novel method, **Adaptive Beta Shrinkage (ABS)**

Summary:

- Better covariance estimation
- Better asset allocation
- More return & less risk

Presentation Overview

1. Introduction and motivation
   - *i. Sample covariance matrix*
   - *ii. Shrinkage- what & why?*
2. Methods
   - *i. Contemporary methods*
   - *ii. Bayesian shrinkage*
   - *iii. Adaptive Beta Shrinkage (ABS)*
3. Backtest results
4. Conclusion



Introduction- sample covariance matrix

N = assets
T = time periods

Sample covariance matrix issues:

- High estimation error => underestimates risk
  - Error increases as N/T increases
- Amplified error when inverted
- Can't invert when N >= T
- Higher trading expenses

Solution? ⟶ **Shrinkage**

Introduction- shrinkage

Shrinkage:

What is it?

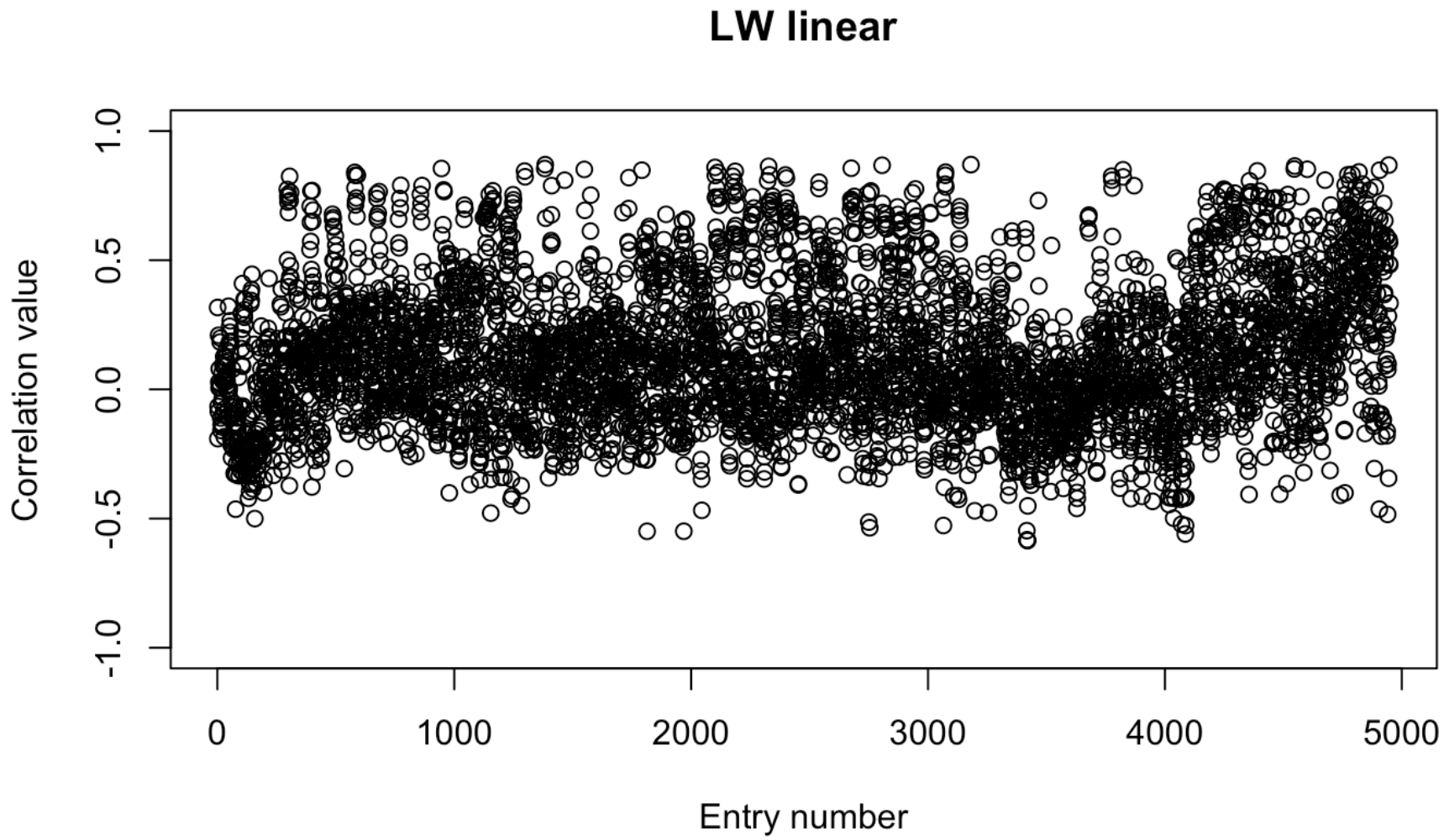
Statistical technique

- combines sample estimate with a stable target
- introduce bias to reduce variance
- $\hat{\Sigma}_{shrunk} = (1-\gamma)\,\hat{\Sigma}_{sample} + \gamma\,\Sigma_{target}$

Why use it

- Reduce **estimation error**
- Fix **dimensionality issues**
- Lower **trading expenses**

Shrinkage visualised



Contemporary methods: Ledoit & Wolf (LW) linear (2004)

$$\hat{\Sigma}_{shrunk} = (1-\gamma)\,\hat{\Sigma}_{sample} + \gamma\,\Sigma_{target}$$

LW linear

Target: 0 correlation

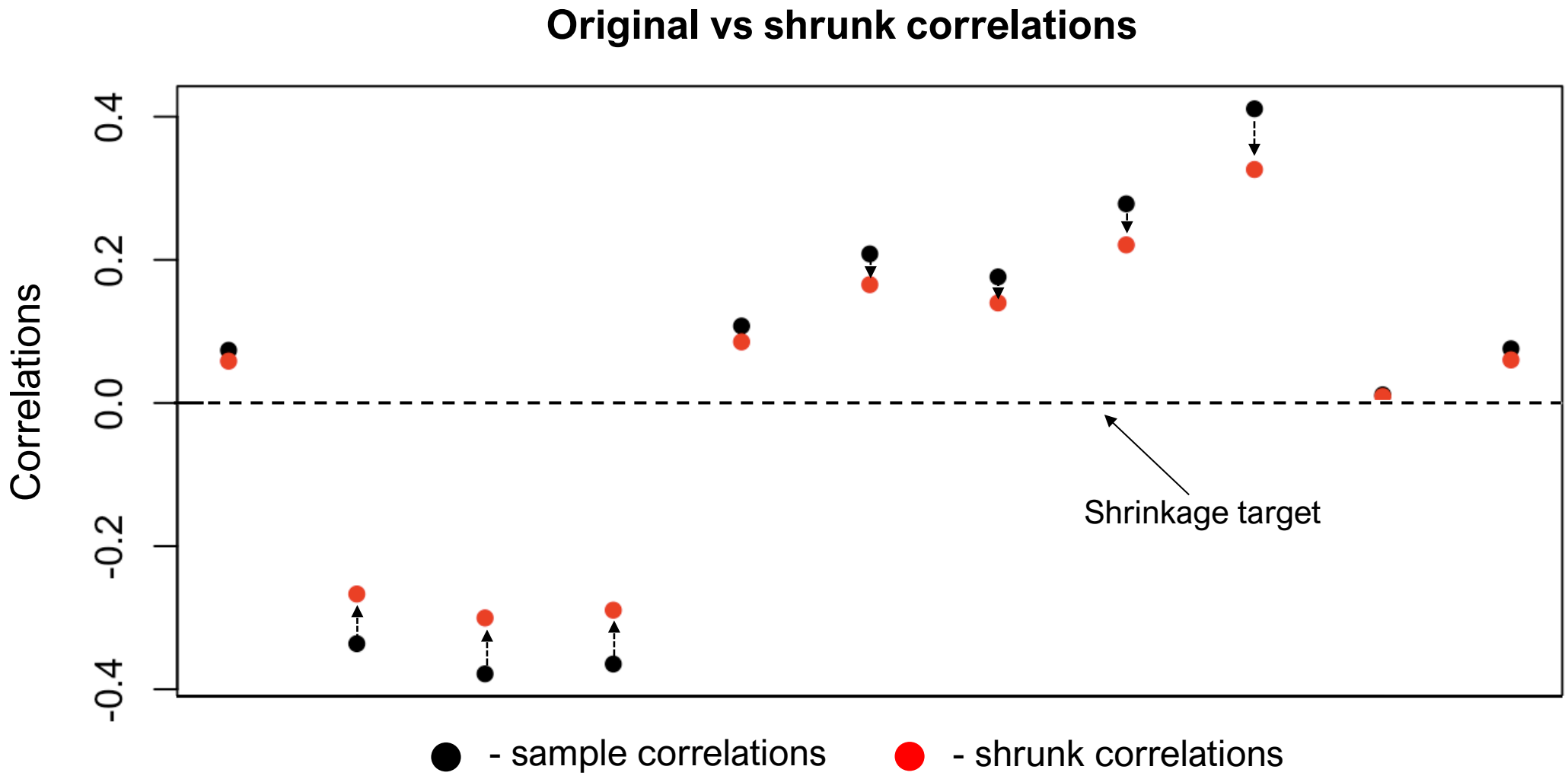
$$\hat{\Sigma}_{target} = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

LW linear

Target: average correlation

$$\hat{\Sigma}_{target} = \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\bar{\rho} \\ \sigma_1\sigma_2\bar{\rho} & \sigma_2^2 \end{bmatrix}$$

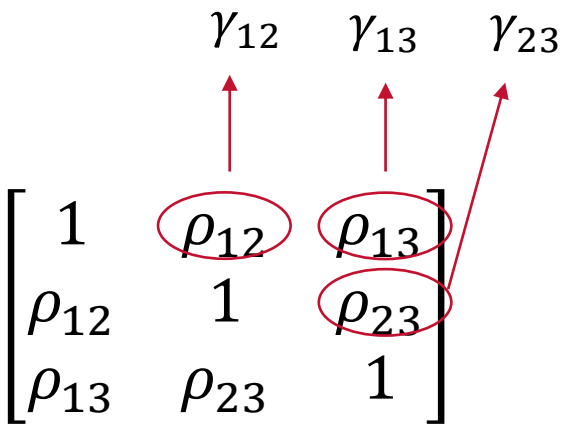
Contemporary methods- LW linear



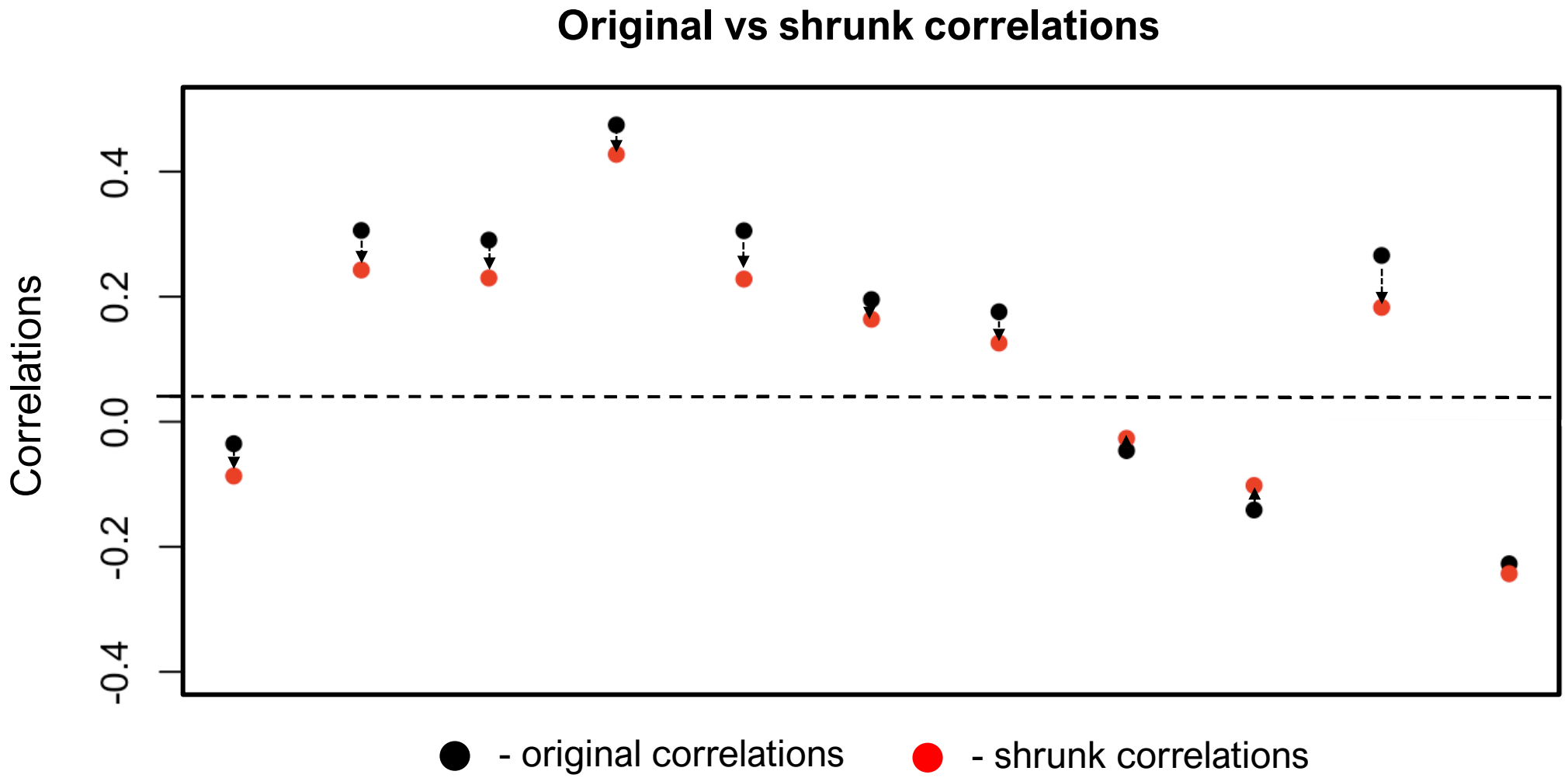
Contemporary methods- LW non-linear (2022)

Non-linear shrinkage:

Different target / shrinkage for **each entry**?



Contemporary methods- LW non-linear (2022) cont.



Contemporary methods- LW non-linear (2022)

Why do non-linear shrinkage?

Improves accuracy

However…

Solve $\frac{N(N-1)}{2} + N$ parameters from T data points:

1. Computationally difficult
2. Invertibility = tricky
3. Overfitting

Difficult for large N!

Contemporary methods- LW non-linear (2022)

$$\hat{\Sigma}_{shrunk} = (1-\gamma)\,\hat{\Sigma}_{sample} + \gamma\Sigma_{target}$$

LW non-linear (QIS)

Why shrink non-linearly + univariately?

→ ***Estimate 1 parameter from T data points***

Spectral decomposition:

$\hat{\Sigma}_{sample} = \mathbf{U\Lambda U'}$

Where $\Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$

Accuracy improvements of NL shrinkage

And avoid main issues of NL:

- Computational
- Non-invertibility
- Overfitting

$\lambda_{1,\,shrunk} = f(\lambda_{1,\,sample})$

$\lambda_{2,\,shrunk} = f(\lambda_{2,\,sample})$

i.e.

- Upsides of NL shrinkage
- None downsides of NL shrinkage

But, can we shrink each entry differently?

Yes! …shrink univariately!

Shrinkage - a Bayesian perspective

posterior

likelihood

prior

$$\hat{\Sigma}_{shrunk} = (1-\gamma)\hat{\Sigma}_{sample} + \gamma\Sigma_{target}$$

prior weight

Univariately- how to shrink eg ρ_{12}?

sample correlation matrix:

likelihood

$$\begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$$

prior = cross-sectional distribution

CorShrink- Dey & Stephens (2018)

1. $\rho_{ij} \rightarrow$ correlation of i and j
   Z_{ij} = arctanh(ρ_{ij}) = $\frac{1}{2} log \left(\frac{1+\rho_{ij}}{1-\rho_{ij}}\right)$
   and $\widehat{Z_{ij}} = \text{arctanh}(\widehat{\rho_{ij}})$

2. Fisher (1921), assuming bivariate normality of i and j:
   Likelihood: $Z_{s,ij} \mid Z_{ij} \sim \text{N}(Z_{ij}, \sigma_{ij})$, with $\sigma_{ij} = \sqrt{\frac{1}{n_{ij}-1} + \frac{2}{(n_{ij}-1)^2}}$

3. Prior: $Z_{ij} \sim D$, D determined empirically

4. Find posterior mean per individual entry

Adaptive Beta Shrinkage (ABS)

Beta:

- Asset's sensitivity to market
- $E(R_i) = R_f + \boldsymbol{\beta} \times (E(R_M) - R_f)$
- Solve via linear regression

Now, consider asset **pairwise betas**:

- $\boldsymbol{\beta_{ij}} = \boldsymbol{\rho_{ij}} \frac{\boldsymbol{\sigma_i}}{\boldsymbol{\sigma_j}}$
- If we standardise the data, $\boldsymbol{\beta_{ij}} = \boldsymbol{\rho_{ij}}$
- So, we can use **regression** to estimate **correlations**

Why shift to regression framework?

- Use **regression techniques**, eg ~~visual estimation~~, weighting
- Use known **beta-adjustment** techniques directly on correlations, eg Vasicek, Blume

ABS (cont.)

Proposed method:

1. Volatility adjustment — *eg divide by EWMA volatilities*

2. Standardise return data → *correlation = beta*

3. Estimate pairwise betas (correlations) via ridge regression + weighting

$$\beta_{xy} = \beta_{lik} = \frac{\boldsymbol{x}'\boldsymbol{V}\boldsymbol{y} + \lambda\beta_{prior}}{\boldsymbol{x}'\boldsymbol{V}\boldsymbol{x} + \lambda}$$

4. Adjust betas with Vasicek (1973) based on estimation error per entry

$$\beta_{post} = \frac{\frac{\beta_{prior}}{\sigma^2_{prior}} + \frac{\beta_{lik}}{\sigma^2_{lik}}}{\frac{1}{\sigma^2_{prior}} + \frac{1}{\sigma^2_{lik}}}$$

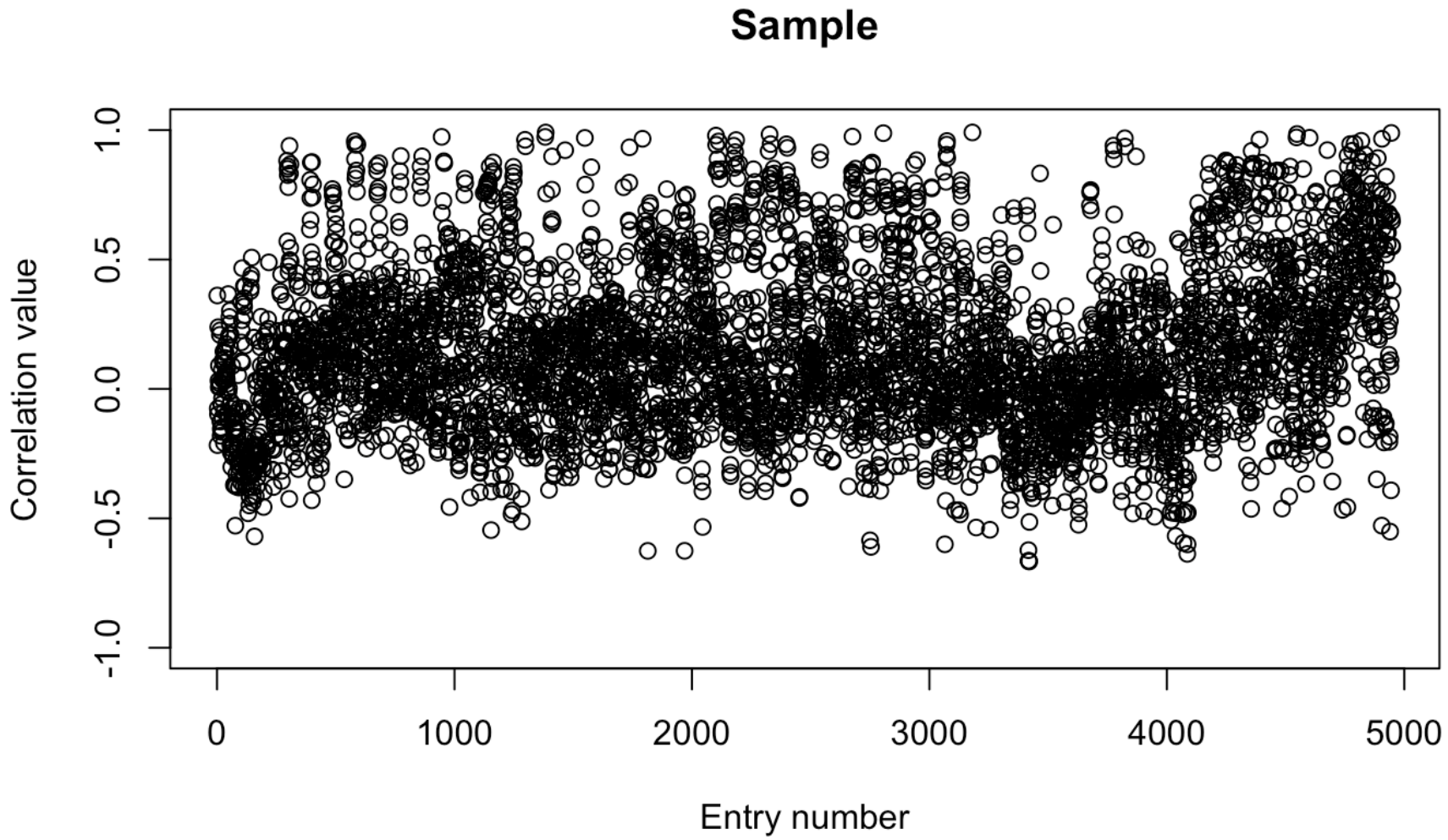
ABS - Summary

Summary:

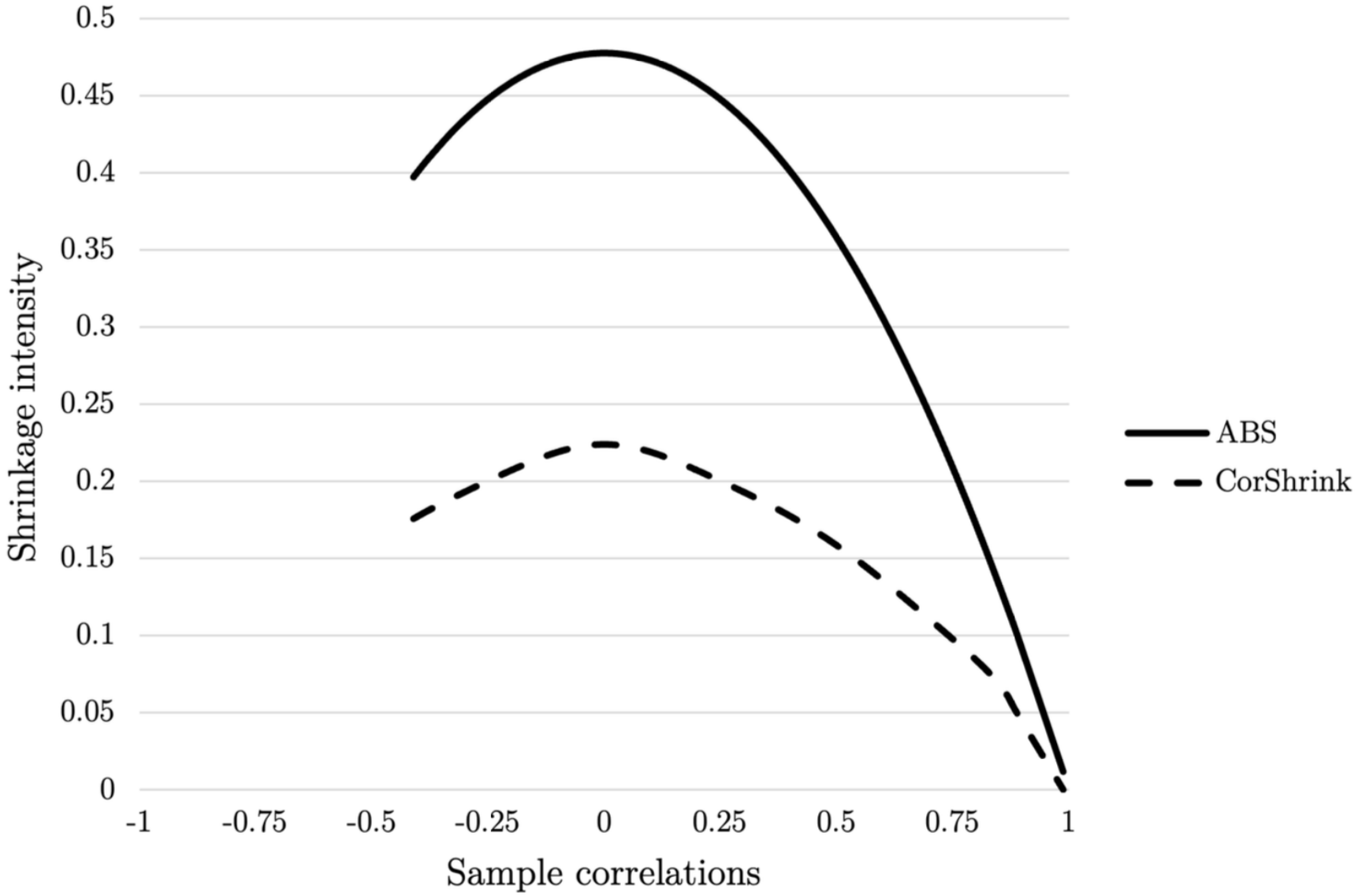
1. **Volatility** adjustment
2. **Standardise** data
3. **Regression** estimation of **beta**
4. **Beta adjustment** method

That's it! Use simple known methods ⟶ non-linear shrinkage

ABS visualised



ABS visualised (cont.)



Backtest results

Methodology:

- Vary **N** size
- **T = 60** months as **in-sample** period
- Estimate covariance matrix with each method
- **1** month **out-of-sample** period
- **Roll forward** by 1 month
- Weights constrained to **0-10%**
- **0.1%** transaction fee

2 Backtests:

- **Minimum Variance** portfolio
- **Max return** with **5% tracking error** to benchmark

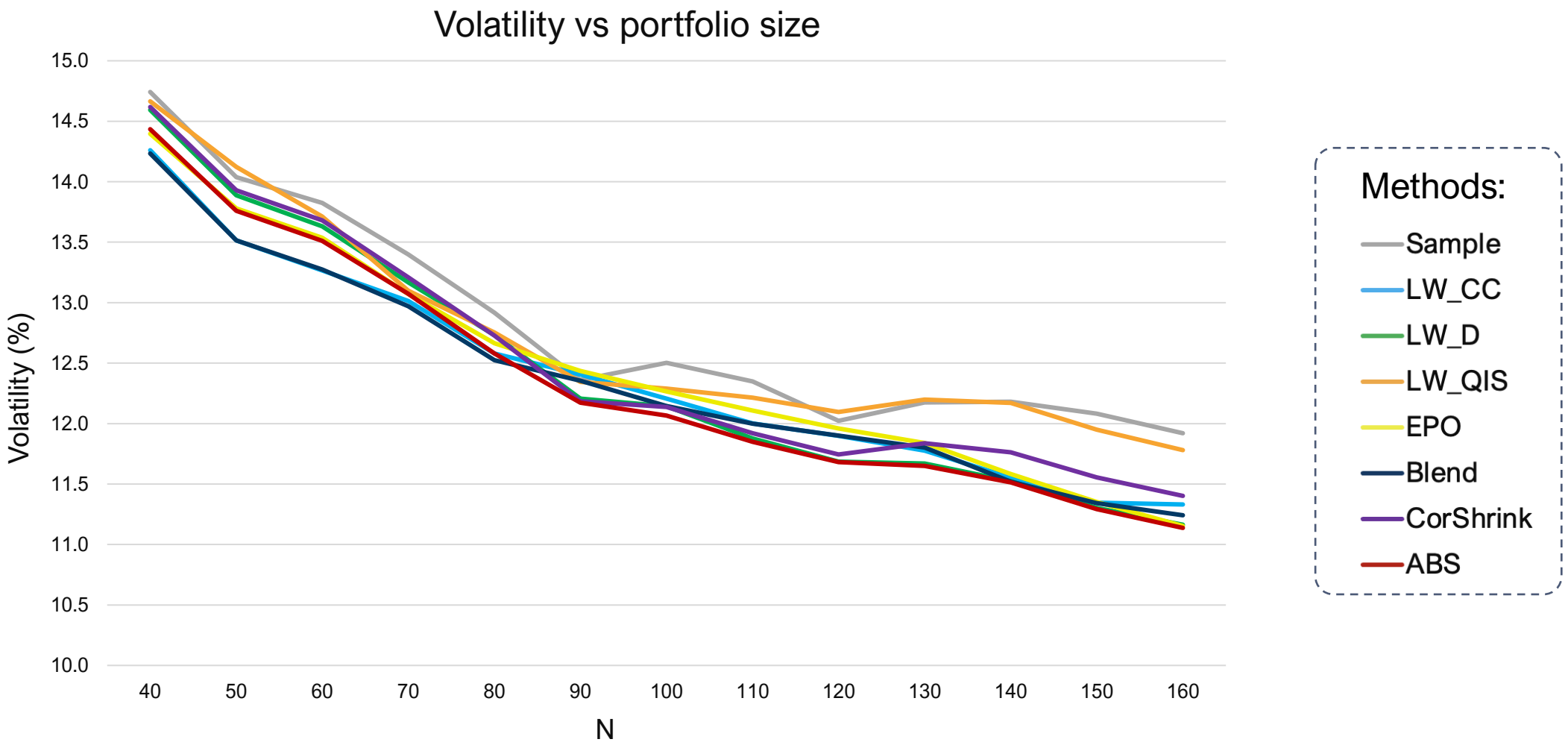
2 Datasets:

- JSE (30 years)
- NASDAQ (20 years)

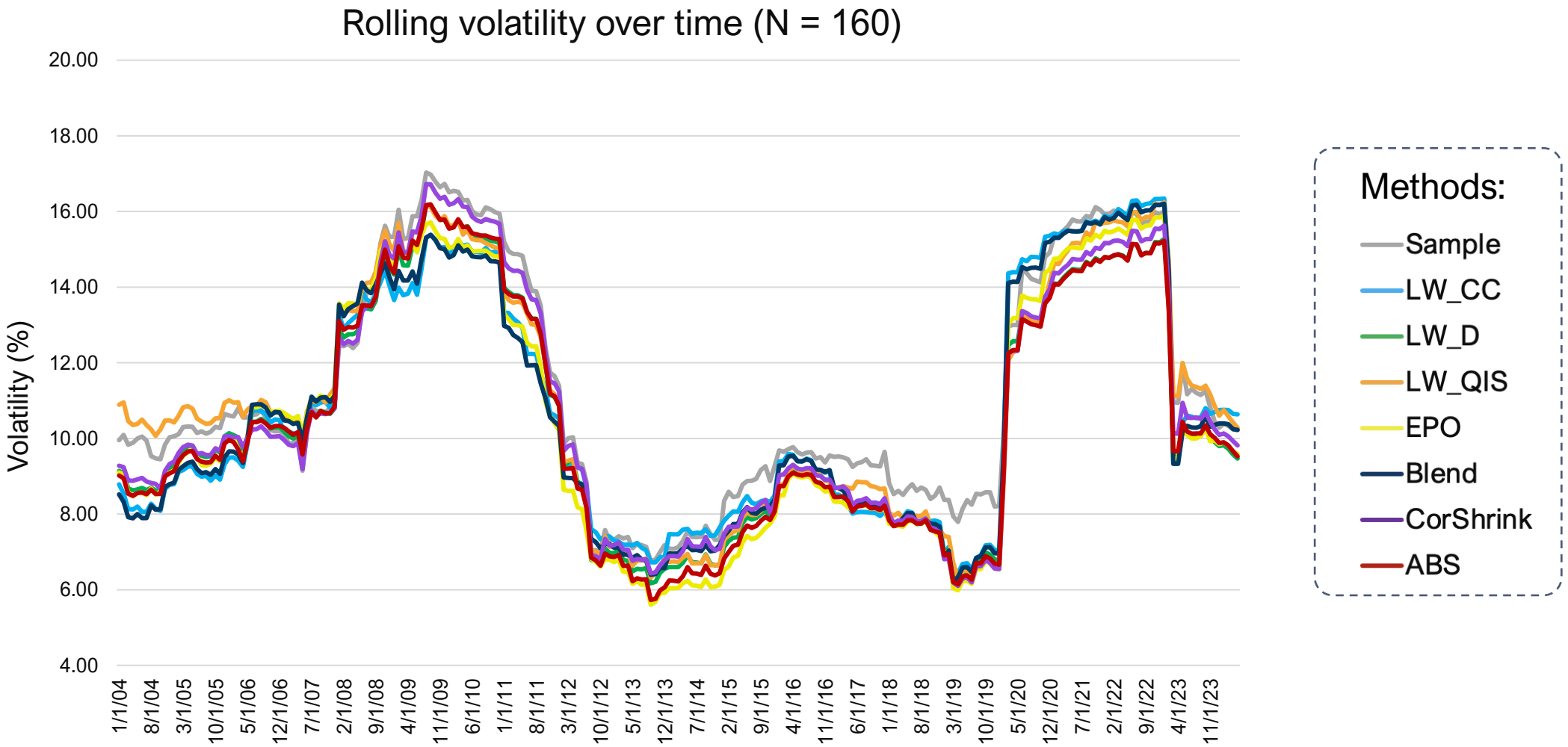
Models:

- LW_CC (2004)
- LW_D (2004)
- LW_QIS (2022)
- EPO (2021)
- Blend (2010)
- CorShrink (2018)
- ABS

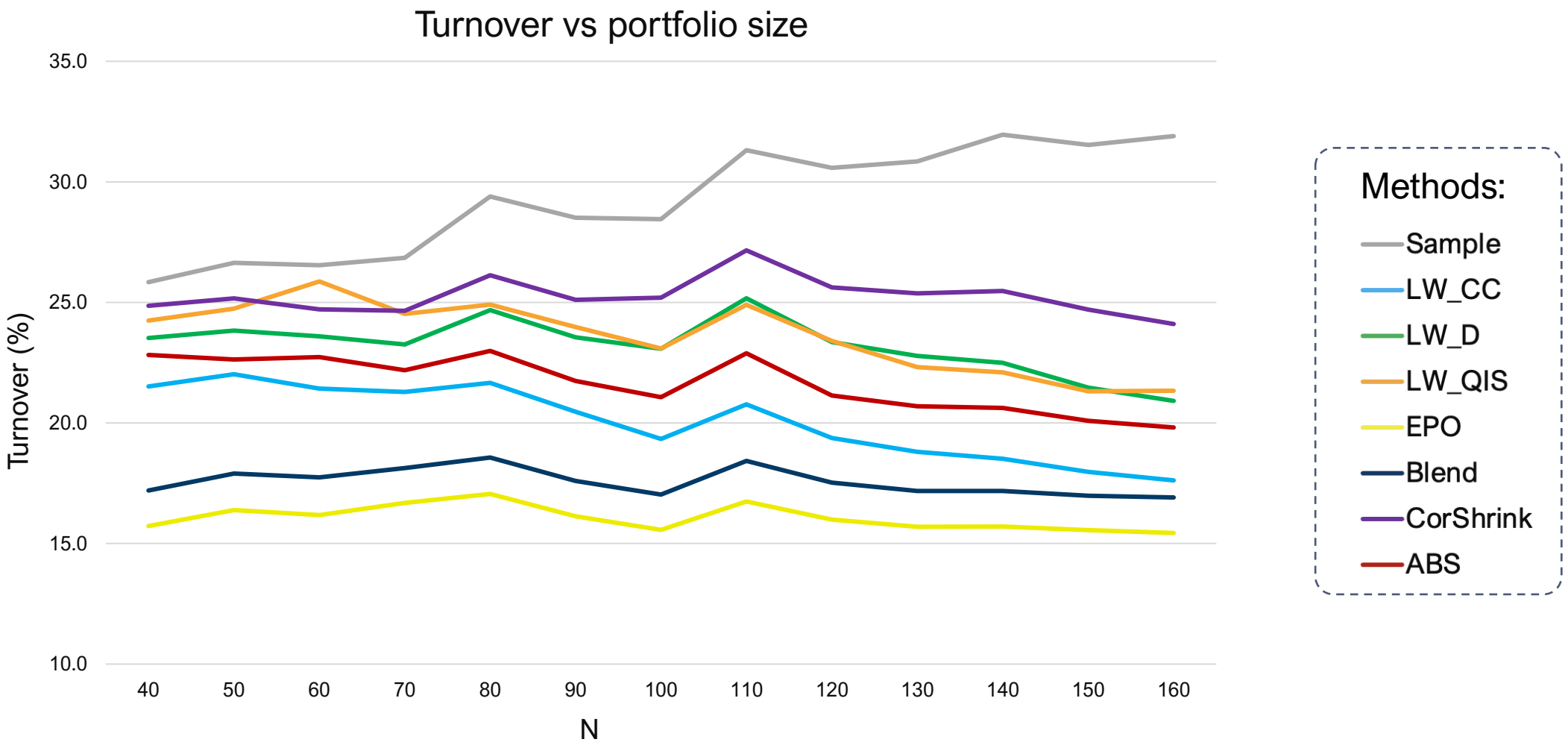
Minimum variance backtest (JSE) - volatility



Minimum variance backtest (JSE) – rolling volatility



Minimum variance backtest (JSE) – turnover

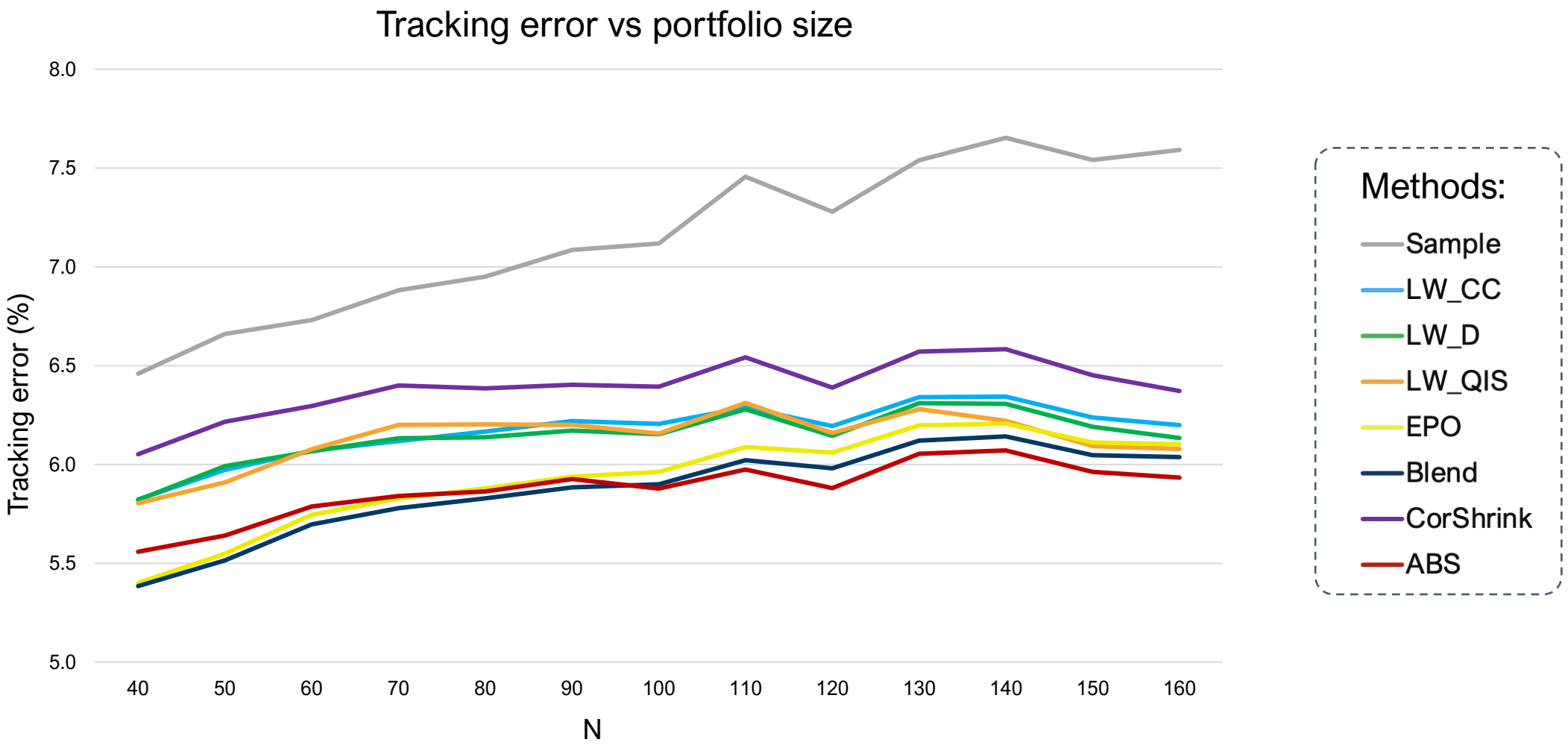


Minimum variance backtest (NASDAQ) – volatility & turnover

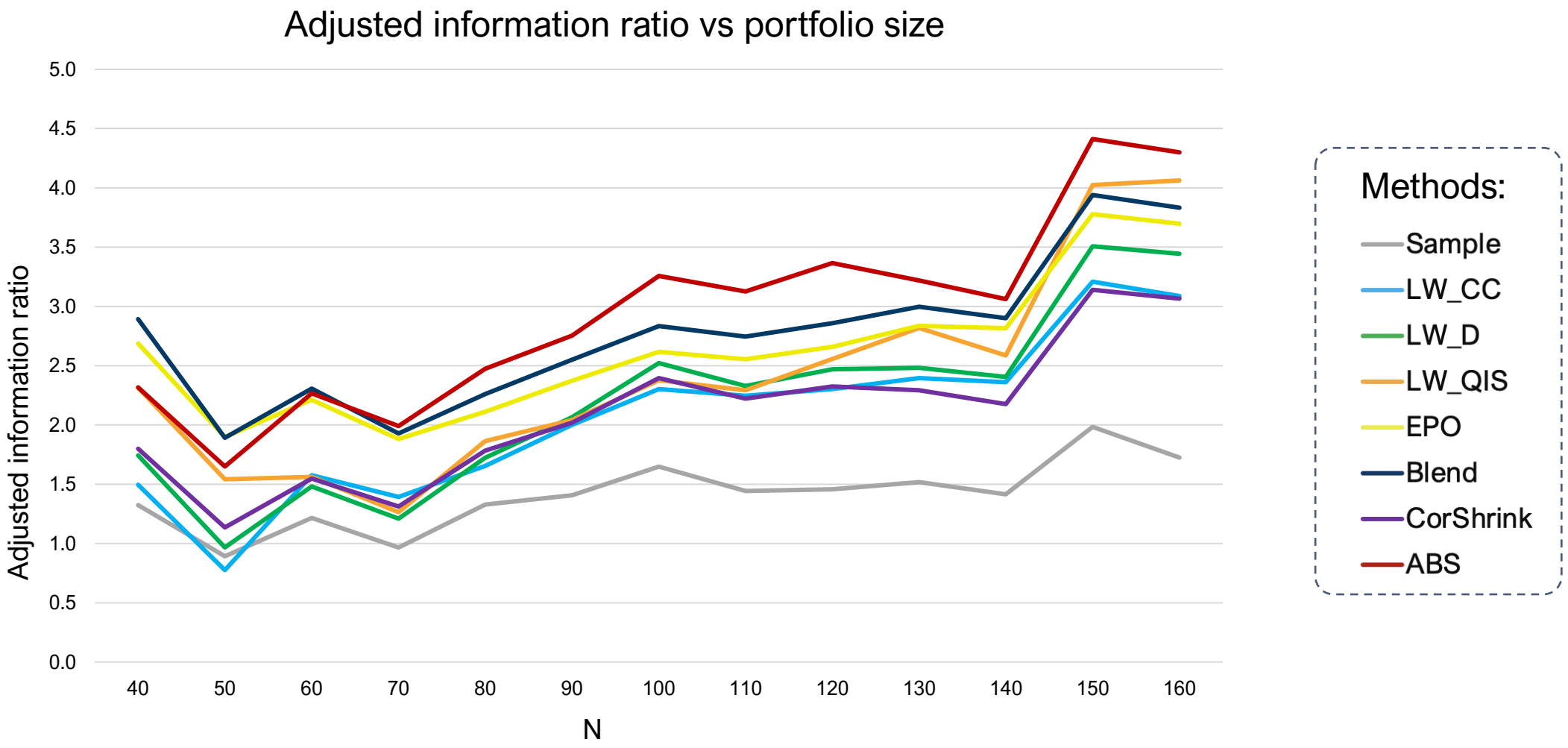
	N = 40		N = 100		N = 160	
	Volatility	Turnover	Volatility	Turnover	Volatility	Turnover
Sample	9.39	20.35	9.92	12.78	12.61	12.24
LW_D	9.06	15.08	**9.48**	12.60	12.51	10.33
LW_CC	8.71	10.99	9.51	11.45	12.47	8.54
LW_QIS	10.49	14.29	9.93	14.64	12.83	13.18
EPO	9.13	**10.75**	**9.48**	9.74	12.67	**8.12**
Blend	**8.63**	10.83	9.93	**9.73**	**12.38**	8.30
CorShrink	9.19	16.77	9.83	13.24	12.55	11.06
ABS	9.13	14.21	9.55	9.86	12.54	9.99

*Best performer for each column shown in **red**

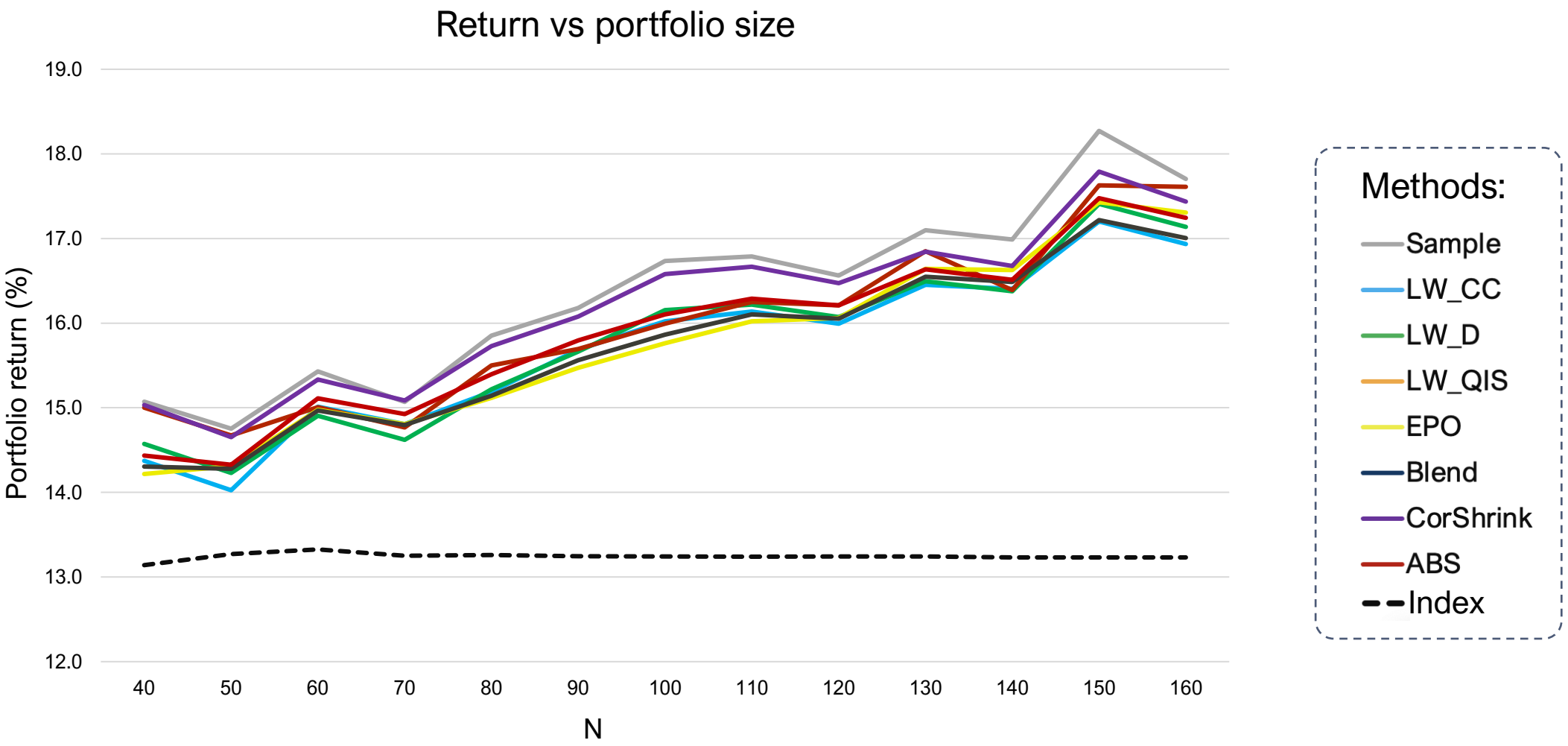
Tracking error backtest (JSE) – tracking error



Tracking error backtest (JSE) – risk-adjusted return



Tracking error backtest (JSE) – returns



Tracking error backtest (NASDAQ)

	N = 40				N = 100				N = 160			
	TE	IR	IR*	$r_p - r_b$	TE	IR	IR*	$r_p - r_b$	TE	IR	IR*	$r_p - r_b$
Sample	7.54	-0.01	-0.04	-0.05	8.69	0.05	0.12	0.49	9.23	0.04	0.09	0.40
LW_D	6.67	0.01	**-0.01**	0.13	7.37	**0.08**	0.25	**0.63**	7.72	**0.09**	0.26	**0.70**
LW_CC	6.80	-0.01	-0.14	-0.02	7.54	0.04	0.11	0.34	7.94	0.04	0.11	0.36
LW_QIS	6.58	0.00	-0.09	0.02	7.16	0.06	0.21	0.48	7.55	0.07	0.22	0.56
EPO	**6.47**	**0.02**	-0.08	**0.16**	7.17	0.07	0.22	0.55	7.54	0.08	0.24	0.62
Blend	6.54	0.00	-0.11	0.07	7.20	0.06	0.17	0.44	7.56	0.06	0.17	0.46
CorShrink	6.94	0.00	-0.04	0.02	7.70	0.07	0.19	0.53	8.07	0.07	0.18	0.57
ABS	6.53	**0.02**	**-0.01**	0.14	**7.14**	**0.08**	**0.26**	0.59	**7.42**	**0.09**	**0.27**	0.67

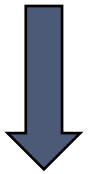
*Best performer for each column shown in **red**

Conclusion

Best method:

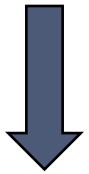
(risk, risk adjusted returns and fees)

Medium – large portfolios:



ABS

Small portfolios:



Blend

- MVO substantially outperforms the index with similar risk levels

Conclusion

Further extentions:

1. Account for non-stationarity
2. More sophisticated expected return estimators
3. Passive index-tracking applications

Thank you!

Questions?

Publication Link:

https://www.tandfonline.com/doi/full/10.1080/10293523.2025.2553255

Contact details:

Email:
henristaaljr@gmail.com

LinkedIn: