Advances in Estimating Factor Correlations

Jose Menchero
Head of Portfolio Analytics Research
jmenchero@bloomberg.net

Antonios Lazanas
Head of Portfolio, Index & ESG Research
alazanas5@bloomberg.net

Outline

- Portfolio Optimization with Noisy Covariance Matrices
  - The “curse of dimensionality”
  - Poor behavior of optimized portfolios
- Correlation Shrinkage
  - Naïve shrinkage (toward zero): beneficial for optimization, but detrimental for risk forecasting
- Advances in Estimating Covariance Matrices
  - Avoiding the pitfalls of the sample correlation and naïve shrinkage
  - Applications to MAC risk models

References:

1. Menchero, Jose and Lei Ji. *Portfolio Optimization with Noisy Covariance Matrices,* Journal of Investment Management (2019)
2. Menchero, Jose and Peng Li. *Correlation Shrinkage: Implications for Risk Forecasting,* Journal of Investment Management (2020)
3. Menchero, Jose and Lei Ji. *Advances in Estimating Covariance Matrices,* Journal of Investment Management (to appear)

Bloomberg

Portfolio Optimization with Noisy Covariance Matrices

The Curse-of-Dimensionality Problem

- Covariance matrices can be very large
  - Russell 3000 (N = 3000 stocks)
  - Bloomberg MAC3 Model (K = 2000+ factors)
- Portfolio construction demands a robust covariance (correlation) matrix
  - Risk model should not identify spurious hedges that fail out-of-sample
- If the number of time periods is less than the number of variables, the sample correlation matrix contains “zero eigenvalues”
  - Leads to spurious prediction of “riskless” portfolios
- This feature makes the sample correlation matrix unsuitable for portfolio construction
- However, the sample correlation is essentially optimal for risk forecasting
- Dual objectives of building a sound risk model:
  - Deviate minimally from the sample correlation to ensure accurate risk forecasts
  - Build a well-conditioned covariance matrix that is reliable for portfolio construction

Bloomberg

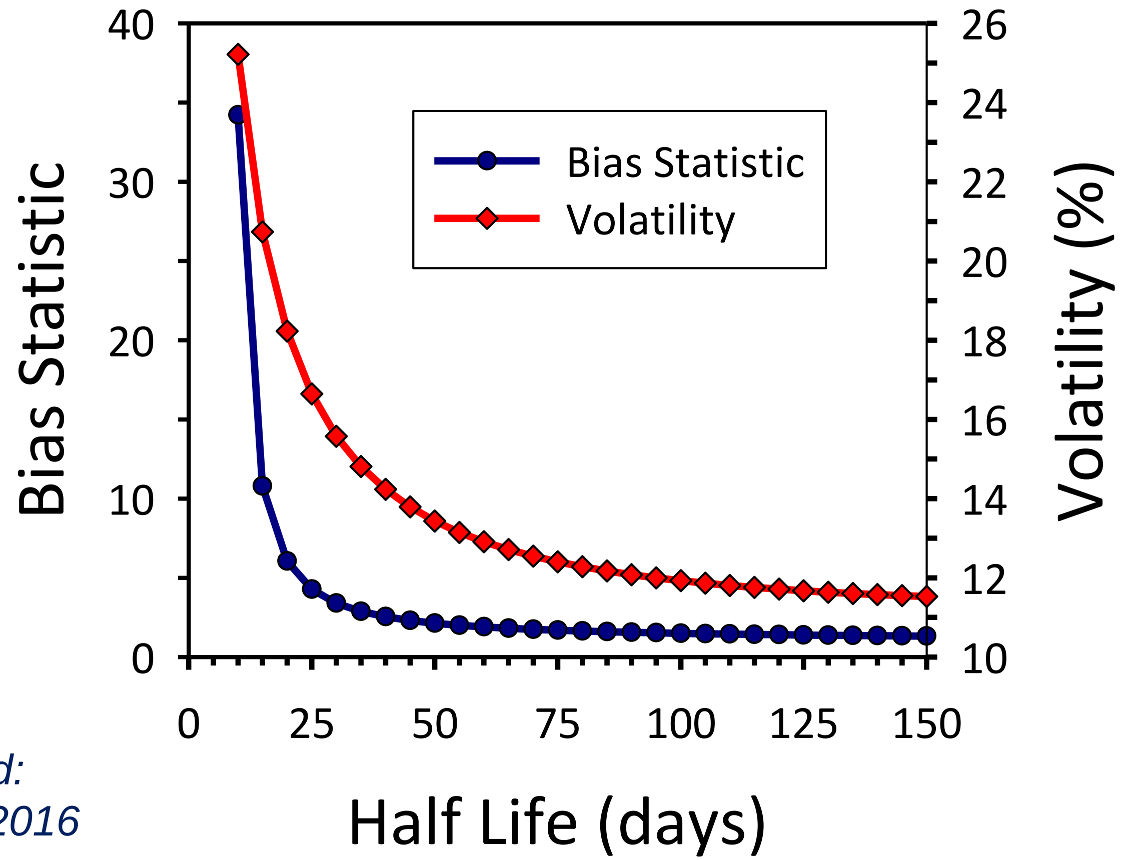
Perils of Noise in the Sample Correlation

- Take the largest 100 US equities as of 31-Mar-2016, with complete daily return history to 13-Jan-1999
- Estimate family of asset covariance matrices using EWMA with a variable HL parameter τ
- Half-life provides a convenient “knob” to control noise level
- Each day, construct the min-vol fully invested portfolio:

$$\mathbf{w}_\tau = \frac{\mathbf{\Omega}_\tau^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Omega}_\tau^{-1}\mathbf{1}}$$

Minimum-volatility fully invested portfolio

B-stat and Volatility



- Bias statistic represents ratio of realized risk to forecast risk
- Biases and out-of-sample volatilities increase dramatically as the covariance matrix becomes more ill-conditioned (noisier)

Out-of-sample test period:
27-Dec-2000 to 31-Mar-2016

Portfolio Optimization (*Ex Ante*)

- Decompose optimal portfolio into alpha and hedge portfolios:

$$\mathbf{w} = \frac{\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}}{\boldsymbol{\alpha}'\boldsymbol{\Omega}^{-1}\boldsymbol{\alpha}} \equiv \boldsymbol{\alpha} + \mathbf{h}$$

Optimal portfolio

$$\mathbf{w} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \\ h_N \end{bmatrix}$$

Alpha Hedge

- Hedge Portfolio is uncorrelated with the optimal portfolio

$$\mathbf{h}'\boldsymbol{\Omega}\mathbf{w} = 0$$

(Property 1)

- Hedge Portfolio has zero alpha

$$\mathbf{h}'\boldsymbol{\alpha} = 0$$

(Property 2)

- Hedge Portfolio is negatively correlated with the alpha portfolio

$$\mathbf{h}'\boldsymbol{\Omega}\boldsymbol{\alpha} < 0$$

(Property 3)

Hedge portfolio reduces portfolio risk without changing the expected return

Bloomberg

Geometry of Optimization (*Ex Ante*)

- Hedge portfolio is uncorrelated with optimal portfolio

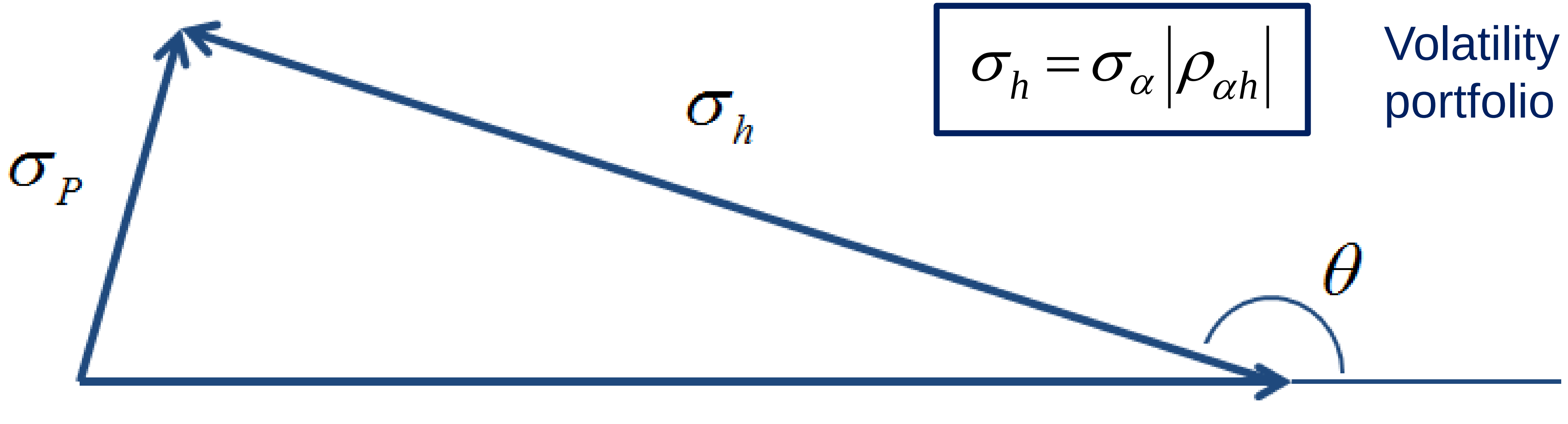
$$\rightarrow \boxed{\sigma_P^2 = \sigma_\alpha^2 - \sigma_h^2}$$

Portfolio Variance

- Let $\rho_{\alpha h}$ denote the predicted correlation between alpha and hedge portfolios
- The magnitude of the correlation determines the quality of the hedge

$$\boxed{\sigma_h = \sigma_\alpha \left|\rho_{\alpha h}\right|}$$

Volatility of hedge portfolio (optimal)



$$\rho_{\alpha h} = \cos(\theta)$$

Bloomberg

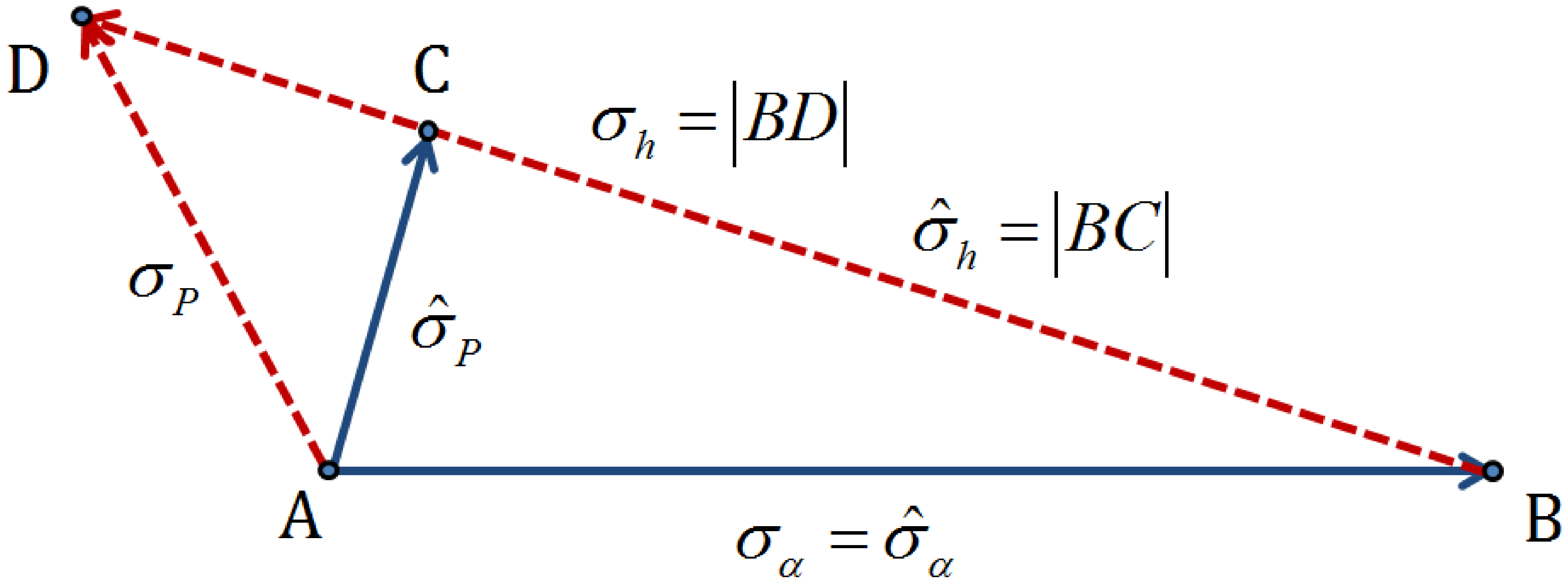
Potential Pitfalls of Optimization

- Optimization always leads to superior *ex ante* performance
- There is no guarantee of improvement *ex post*
- Estimation error within the covariance matrix represents a potential pitfall in portfolio optimization
- Estimation error in the volatility:
  - Risk models may underestimate the volatility of the hedge portfolio
- Estimation error in the correlation:
  - Risk models may “paint an overly rosy picture” of the correlation between the alpha and the hedge portfolios
- Estimation error gives rise to several detrimental effects:
  - Underestimation of risk of optimized portfolios
  - Higher out-of-sample volatility of optimized portfolios
  - Positive realized correlation between optimized and hedge portfolios

Bloomberg

Estimation Error in the Volatility

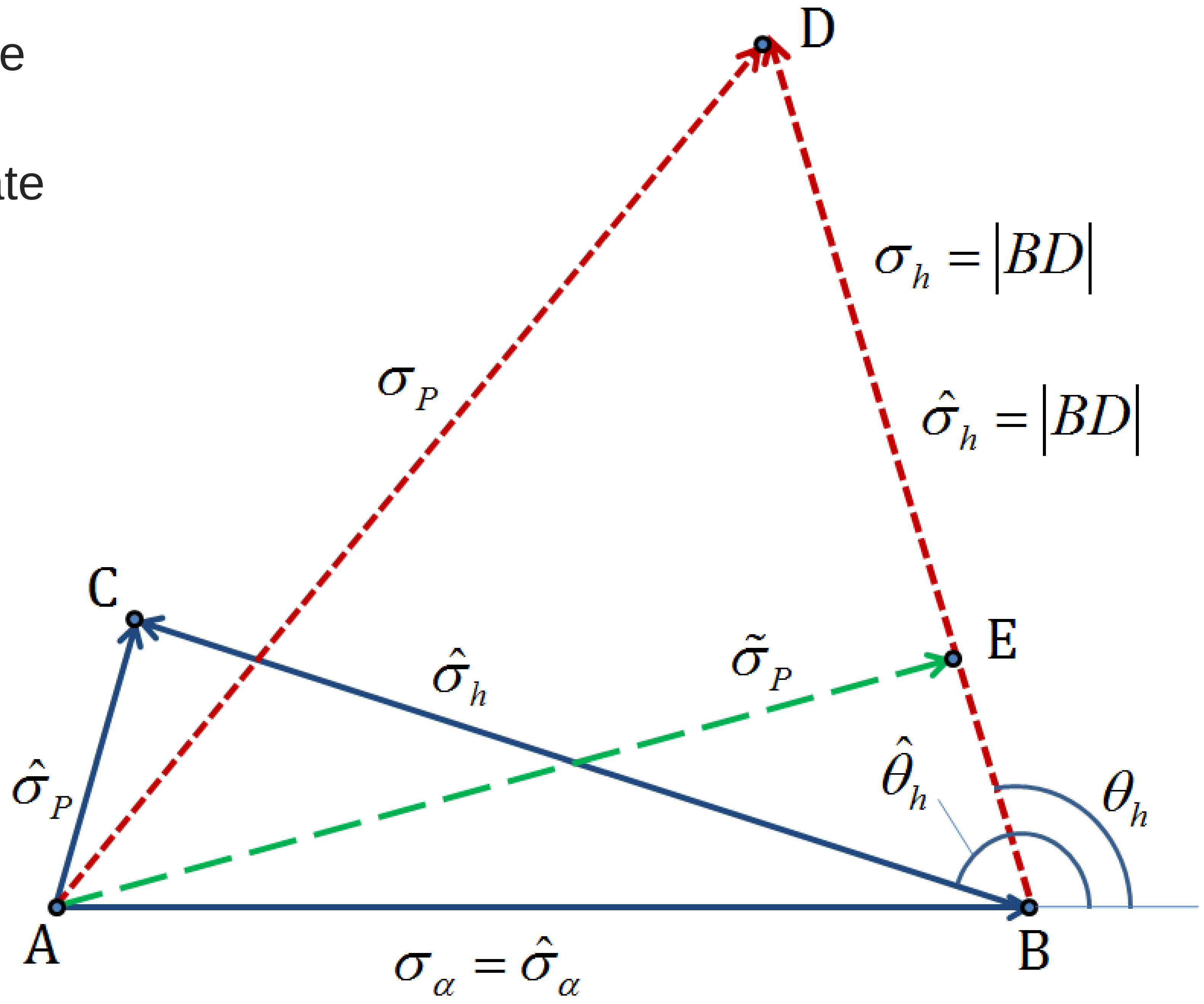
- Suppose we correctly estimate the correlation between hedge/alpha portfolios, but we under-estimate volatility of hedge portfolio
- Side effects:
  - Under-estimation of risk of optimal portfolio
  - Inefficient allocation of risk budget (hedge portfolio adds risk but no return)
  - Increased out-of-sample volatility of optimal portfolio



Bloomberg

Estimation Error in the Correlation

- Now suppose that we correctly estimate the volatility of the hedge portfolio
- However, suppose that we over-estimate magnitude of correlation
- Side effects
  - Under-estimation of portfolio risk
  - Inefficient allocation of risk budget
  - Increased out-of-sample volatility
- We think we hold Portfolio C
- In reality, we hold Portfolio D
- Portfolio E is true optimal portfolio



Empirical Study

- Take the largest 100 US equities as of 31-Mar-2016, with complete daily return history back to 13-Jan-1999
- Estimate a family of asset covariance matrices using EWMA with variable HL parameter τ
- Each day, construct the minimum-volatility portfolio with unit weight in a particular stock:

$$\mathbf{w}_{n\tau} = \frac{\mathbf{\Omega}_\tau^{-1}\boldsymbol{\delta}_n}{\boldsymbol{\delta}_n' \mathbf{\Omega}_\tau^{-1}\boldsymbol{\delta}_n}$$

Optimal portfolio (stock *n*)

- We have a family of 100 portfolios for each HL parameter
- Each portfolio is 100 percent long a particular stock, and hedges the risk by shorting the other 99 stocks
- Decompose 100 portfolios into alpha and hedge portfolios
- Noise may induce biases in predicted volatilities and correlations

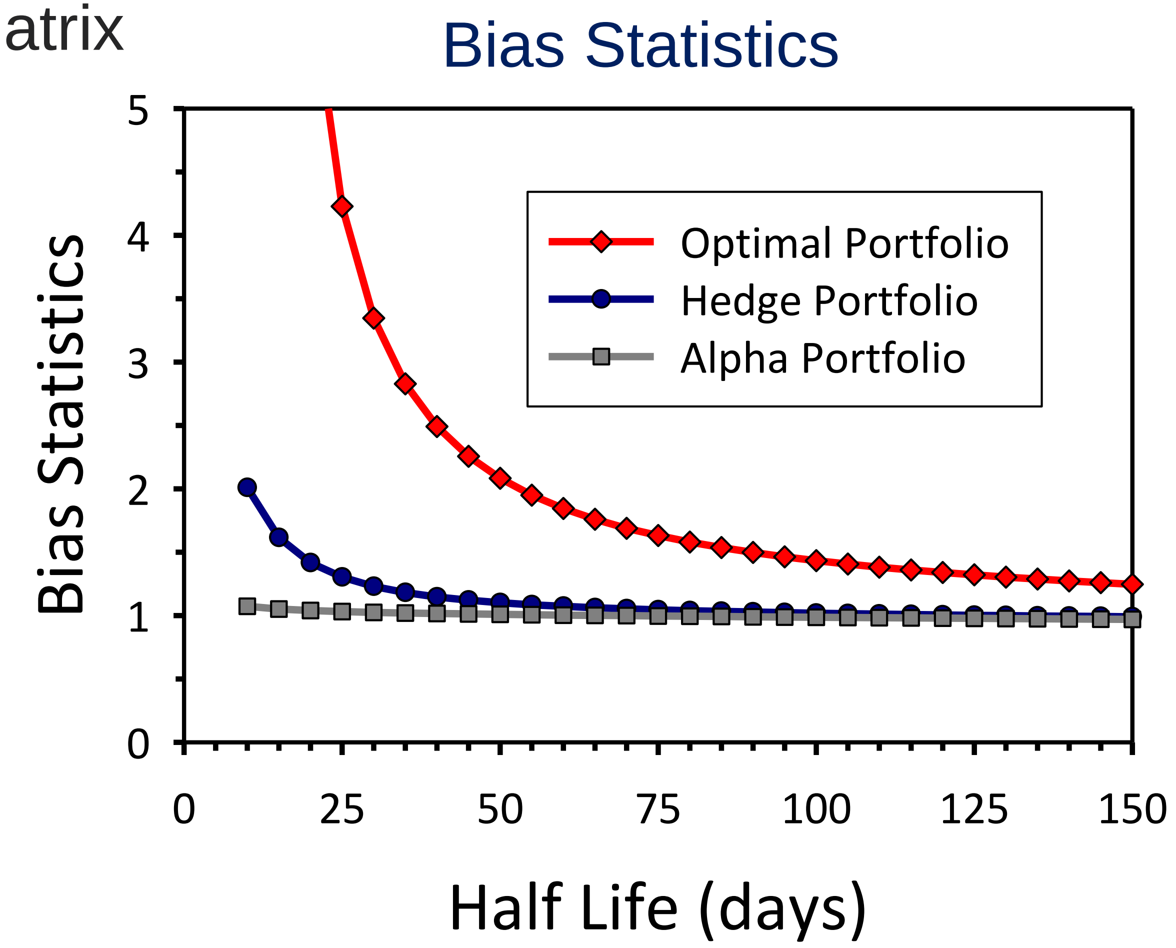
Menchero, Jose and Lei Ji. *Portfolio Optimization with Noisy Covariance Matrices,* Journal of Investment Management (2019)

Bloomberg

Biases in Portfolio Volatility

- Compute mean bias statistics for each of 100 alpha, hedge, and optimal portfolios
- Alpha portfolio is fully invested in a single stock
  - Alpha portfolio is independent of the covariance matrix
  - Covariance matrix makes unbiased forecasts for alpha portfolio
- Hedge/optimal portfolios depend on covariance matrix
  - Hedge portfolio risk is largely unbiased, except for very short HL
  - Optimal portfolio is significantly under-forecast across the full range of HL parameters

Biases in correlations must be responsible for under-forecasting the volatility of optimal portfolio

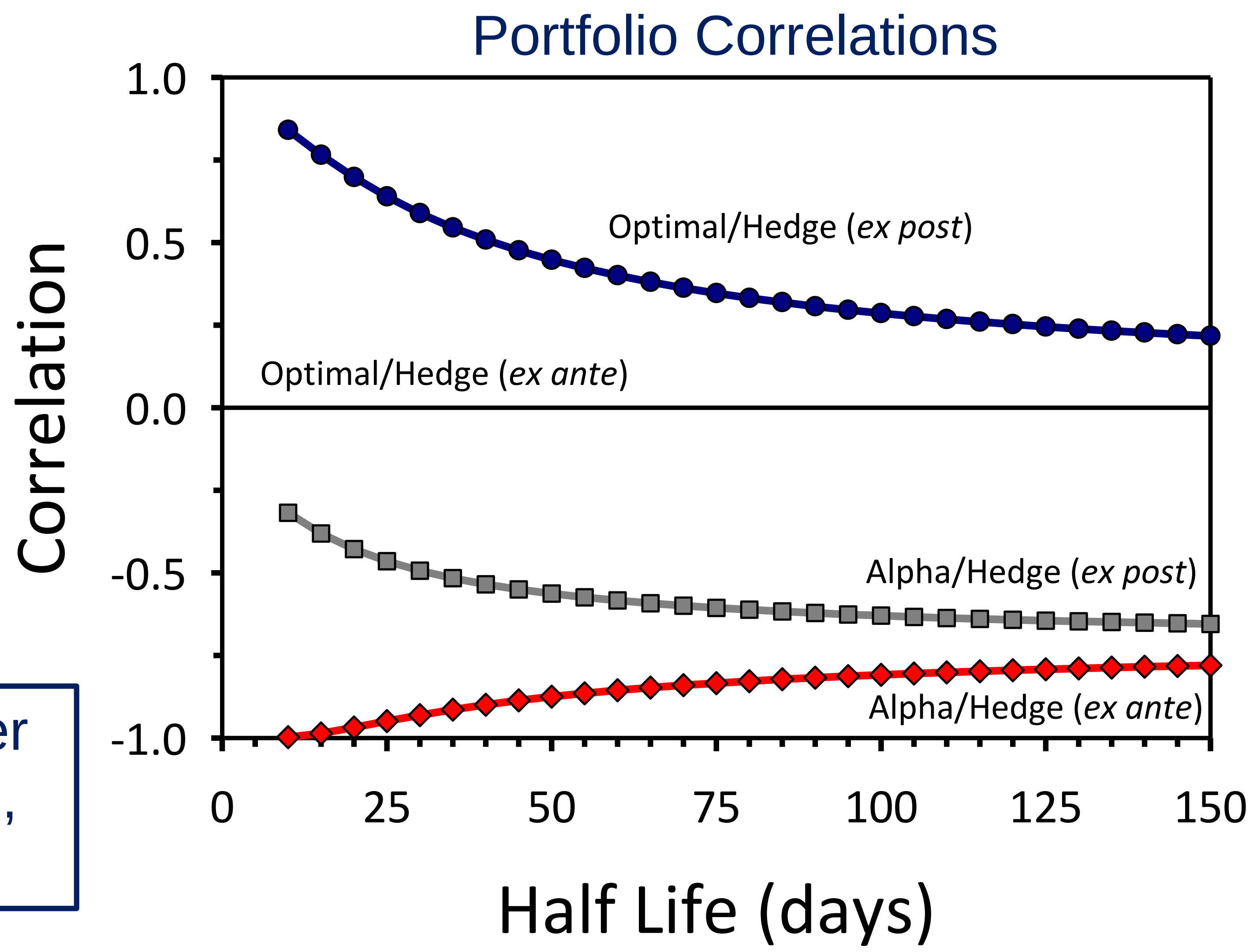


Bloomberg

Biases in Correlations

- Compute mean portfolio correlations both *ex ante* and *ex post*
- Optimal/hedge correlation:
  - *Ex ante*, the correlation is exactly zero
  - *Ex post*, it is positive, indicating inefficient allocation of the risk budget
  - Problem is exacerbated for short HL
- Alpha/hedge correlation:
  - The hedge always appears better *ex ante* than it turns out to be *ex post*
  - Gap between *ex ante* and *ex post* grows larger for short HL parameters

The noisier the correlations, the better the optimal portfolio appears *ex ante*, but the worse it becomes in reality



Bloomberg

Alpha/Hedge Portfolios and the Efficient Frontier

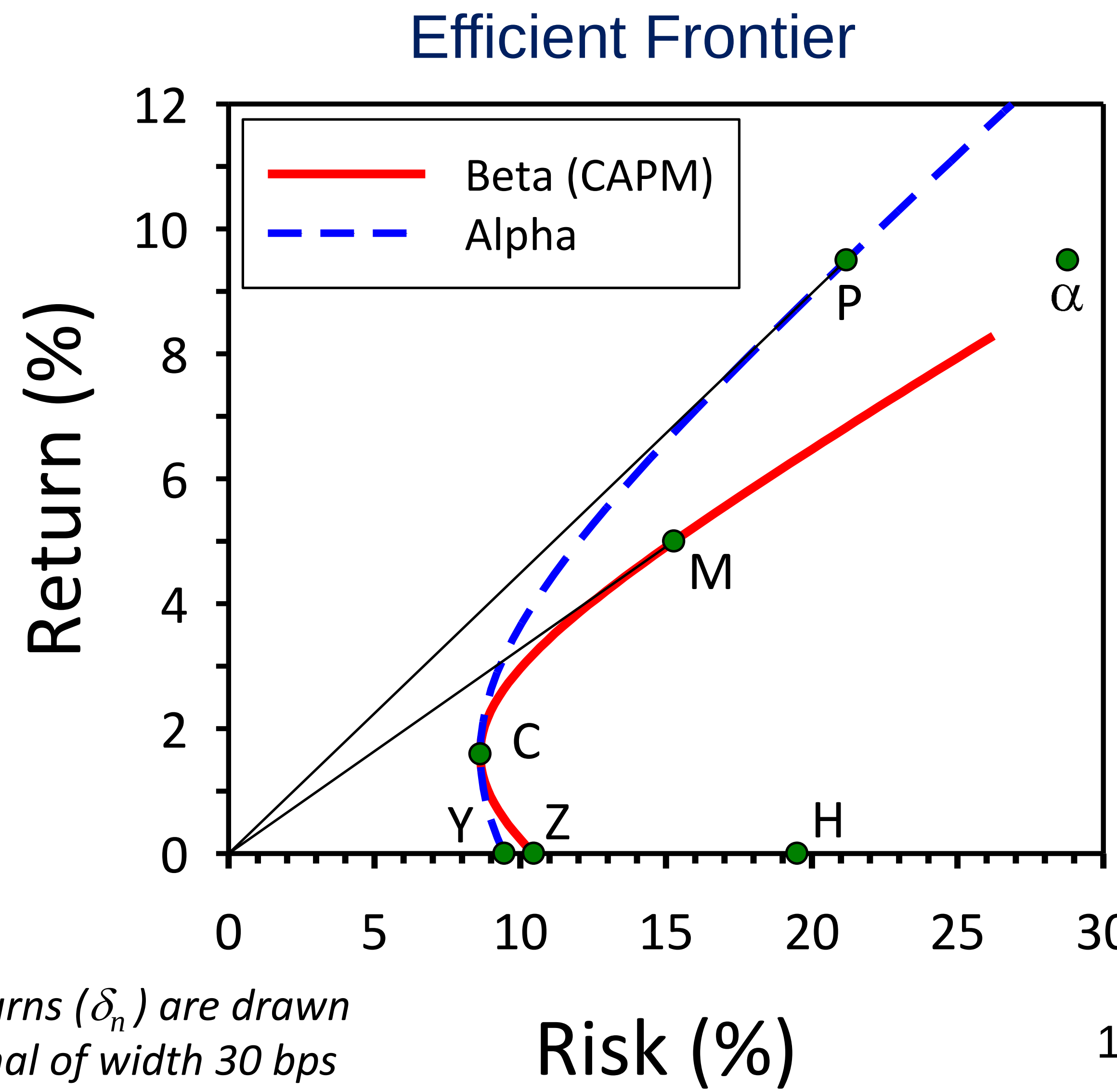
- Form the minimum-volatility fully invested portfolio with fixed expected return
- Universe is largest 100 stocks with covariance matrix using 150-day HL on 31-Mar-2016
- Red line is the efficient frontier assuming CAPM:

$$E\left[r_n\right]=\beta_n E\left[R_M\right]$$

- Blue line is the efficient frontier with alphas (δ_n):

$$E\left[r_n\right]=\beta_n E\left[R_M\right]+\delta_n$$

- C is the min-vol fully invested portfolio
- Y/Z are the zero-beta portfolios (relative to P/M)
- P/M are the efficient portfolios using respective return assumptions
- α is the alpha portfolio (177% long)
- H is the hedge portfolio (77% short)



Note: exceptional returns (δ_n) are drawn from a standard normal of width 30 bps

Constructing the Efficient Frontier

- Portfolio optimization with equality constraints:

  Minimize: $\mathbf{w}'\boldsymbol{\Omega}\mathbf{w}$ Subject to: $\mathbf{A}\mathbf{w} = \mathbf{b}$

- Solution obtained using Lagrange multipliers:

$$\mathbf{w} = \boldsymbol{\Omega}^{-1}\mathbf{A}'\left(\mathbf{A}\boldsymbol{\Omega}^{-1}\mathbf{A}'\right)^{-1}\mathbf{b}$$

Optimal Portfolio

- Target return and full-investment constraints:

$$\left.\begin{array}{ll} \boldsymbol{\alpha}'\mathbf{w} = \mu_P & \text{(Target return)} \\ \mathbf{1}'\mathbf{w} = 1 & \text{(Full investment)} \end{array}\right\} \longrightarrow \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_N \\ 1 & 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} = \begin{bmatrix} \mu_P \\ 1 \end{bmatrix}$$

- Terminology: alphas here represent expected stock returns

$$\alpha_n \equiv E\left[r_n\right] = \beta_n E\left[R_M\right] + \delta_n$$

The δ_n represent exceptional returns, commonly referred to as "alphas"

Bloomberg

Correlation Shrinkage

Shrinking Correlations Toward Zero

- Factorize the sample covariance matrix into product of a diagonal volatility matrix $\hat{\mathbf{V}}$ and a correlation matrix $\hat{\mathbf{C}}$
  - Typically estimate volatilities with short HL parameter (responsiveness)
  - Typically estimate correlations with relatively long HL parameter

$$\boxed{\hat{\mathbf{\Omega}} = \hat{\mathbf{V}}\hat{\mathbf{C}}\hat{\mathbf{V}}}$$ Sample Covariance Matrix

- Shrinking the sample correlation toward zero

$$\tilde{\mathbf{C}}_{\lambda} = (1-\lambda)\hat{\mathbf{C}} + \lambda\mathbf{I} \rightarrow \boxed{\tilde{\mathbf{\Omega}}_{\lambda} = \hat{\mathbf{V}}\tilde{\mathbf{C}}_{\lambda}\hat{\mathbf{V}}}$$ Shrunk Covariance Matrix

- Benefits of shrinking
  - Reduced sampling error
  - May be beneficial for portfolio construction
- Pitfalls of shrinking
  - May produce biased forecasts

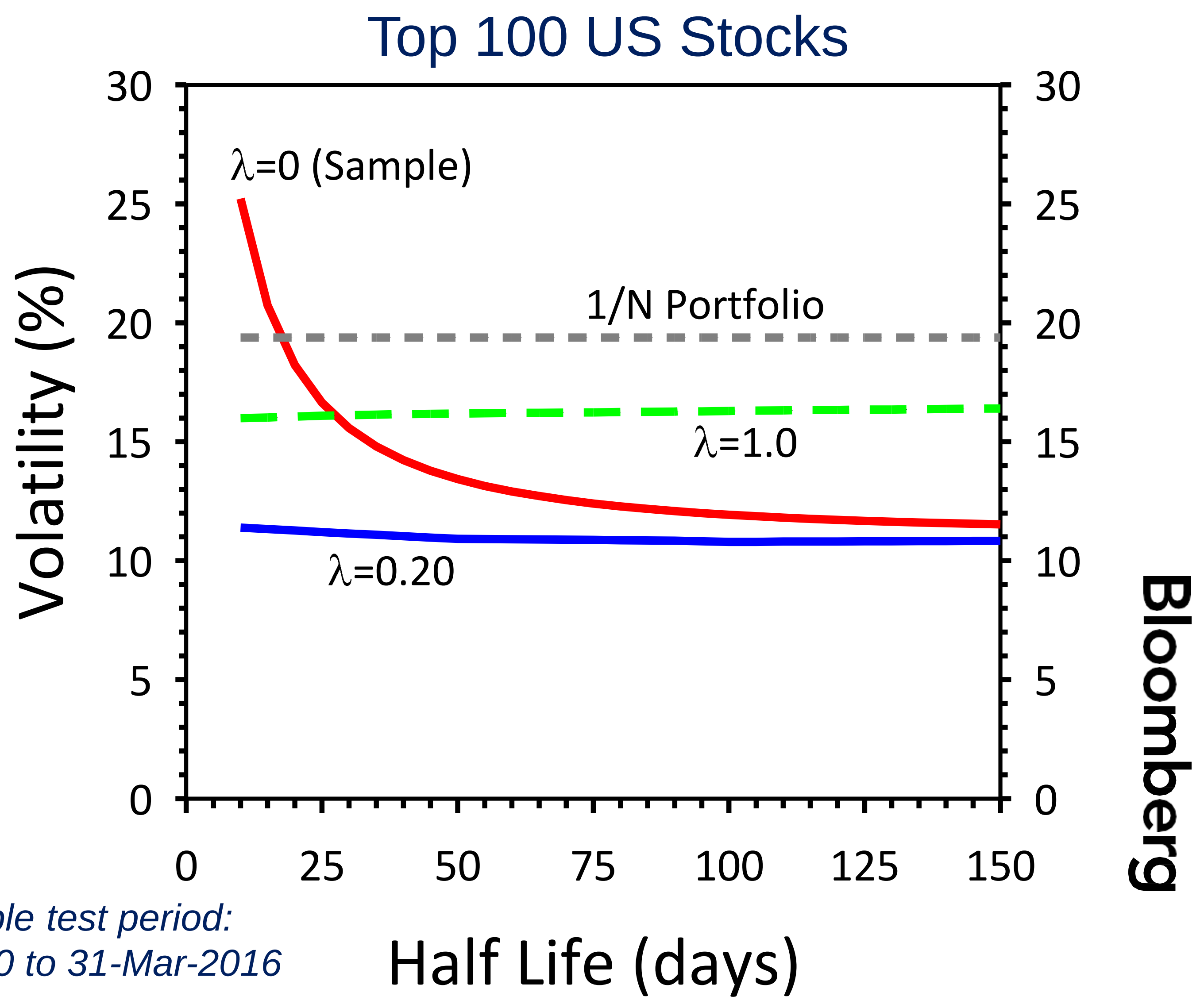
Bloomberg

Shrinkage and Portfolio Optimization

- Construct minimum-volatility fully invested portfolio four different ways:
  - 1/N portfolio (equal weighted, no optimization)
  - Zero correlation (λ=1.0)
  - Sample correlation (λ=0)
  - Partial shrinkage (λ=0.2)

Results:

- 1/N portfolio performed worst, except for very short HL
- Zero correlation amounts to inverse variance weights
- Sample correlation performed worst for very short HL (due to noise)
- Shrinkage of 20% had lowest volatility for any value of HL parameter



Out-of-sample test period:
27-Dec-2000 to 31-Mar-2016

Pitfalls of Naïve Shrinkage

- Consider two-asset portfolios

$$R_P = f_1 - f_2$$ (Go long Asset 1, short Asset 2)

- Suppose assets have equal volatility ($\sigma = 1$) and correlation ρ

$$\longrightarrow \boxed{\sigma_P^2 = 2(1-\rho)}$$ True portfolio variance

- Portfolio variance is estimated over τ periods

$$\hat{\rho}_\lambda = (1-\lambda)\hat{\rho} \quad \rightarrow \quad \boxed{\hat{\sigma}_P^2 = \hat{\sigma}_1^2 + \hat{\sigma}_2^2 - 2\hat{\rho}_\lambda \hat{\sigma}_1 \hat{\sigma}_2}$$ Estimated variance

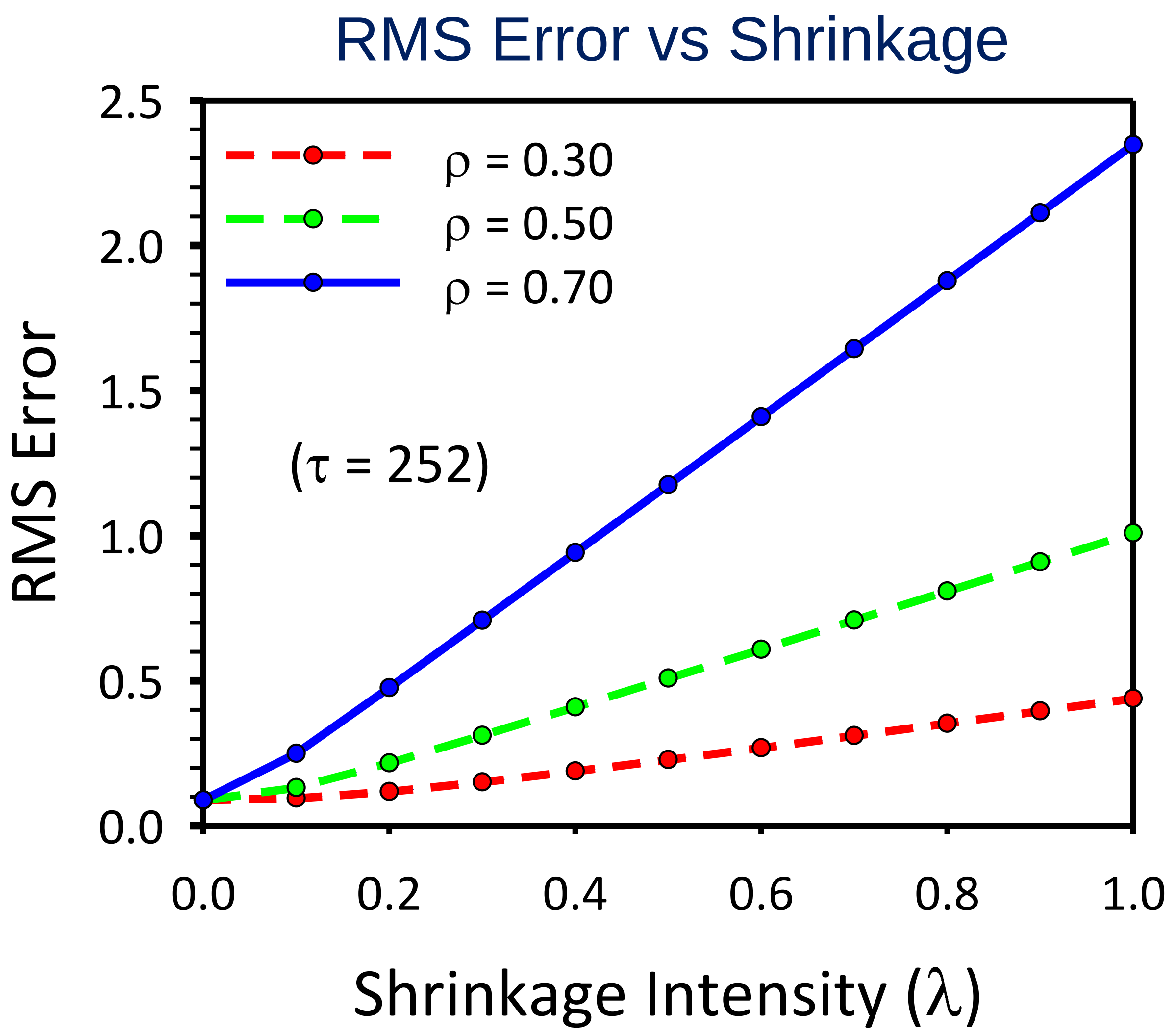
- Error depends on three parameters: (1) number of periods τ, (2) true correlation ρ, and (3) shrinkage intensity λ

$$\delta_{\tau\rho\lambda} = \frac{\hat{\sigma}_P^2 - \sigma_P^2}{\sigma_P^2} \quad \rightarrow \quad \boxed{\varepsilon_{\tau\rho\lambda} = \sqrt{E\left[\delta_{\tau\rho\lambda}^2\right]}}$$ RMS error

Bloomberg

Sample Correlation is Nearly Optimal

- Plot RMS error versus shrinkage intensity
  - Consider three values for true correlation (0.30, 0.50, 0.70)
  - Consider a 252-day look-back window for estimation
- Zero shrinkage appears optimal
  - Actually, finite shrinkage always reduces RMS error
  - Optimal shrinkage intensity is very close to zero
  - The error reduction is so miniscule that it is not visible to the naked eye
  - Shrinkage may induce large errors in risk forecasts
  - Forecasting error is exacerbated as the correlation and/or shrinkage intensity increase
- Effectively, the sample correlation is *optimal* for predicting risk (for non-optimized portfolios)



Menchero, Jose and Peng Li. *Correlation Shrinkage: Implications for Risk Forecasting,* Journal of Investment Management (2020)

Advances in Estimating Factor Correlations

Bloomberg

Estimating Factor Correlation Matrices

- Sample correlation matrix
  - Based on “textbook” definition (i.e., intuitive and transparent)
  - Produces accurate risk forecasts (except for optimized portfolios)
  - Possesses “Achilles heel” (not reliable for portfolio construction)
- Principal Component Analysis (PCA)
  - Statistical technique to extract global factors from the data
  - Assume a small number of global factors (principal components) fully capture correlations of local factors (i.e., uncorrelated residuals)
- Time-series Approach
  - Specify “global” factor returns to explain “local” factor correlations
  - Estimate loadings (exposures) of local factors by time-series regression
  - Local correlations derived from loadings and global covariance matrix
  - May result in biased correlations if important global factors are omitted
  - Correlations can deviate greatly from the sample correlation

Bloomberg

Blended Correlation Matrices

- Ledoit and Wolf (2003) showed that blending the sample covariance matrix with a covariance matrix from a one-factor model produced optimized fully invested portfolios with lower out-of-sample volatility than either model individually
- Blend sample correlation (using weight w) with PCA correlation using J principal components derived from K local factors
- Specify number of PCA factors by parameter μ, where $J = \mu K$
- Two-parameter model for correlation matrix:

$$\mathbf{C}_B(\mu, w) = w\mathbf{C}_0 + (1-w)\mathbf{C}_P(\mu)$$

Blended Matrix

- Optimal blending parameters are determined empirically
- Technique represents the MAC3 Bloomberg methodology

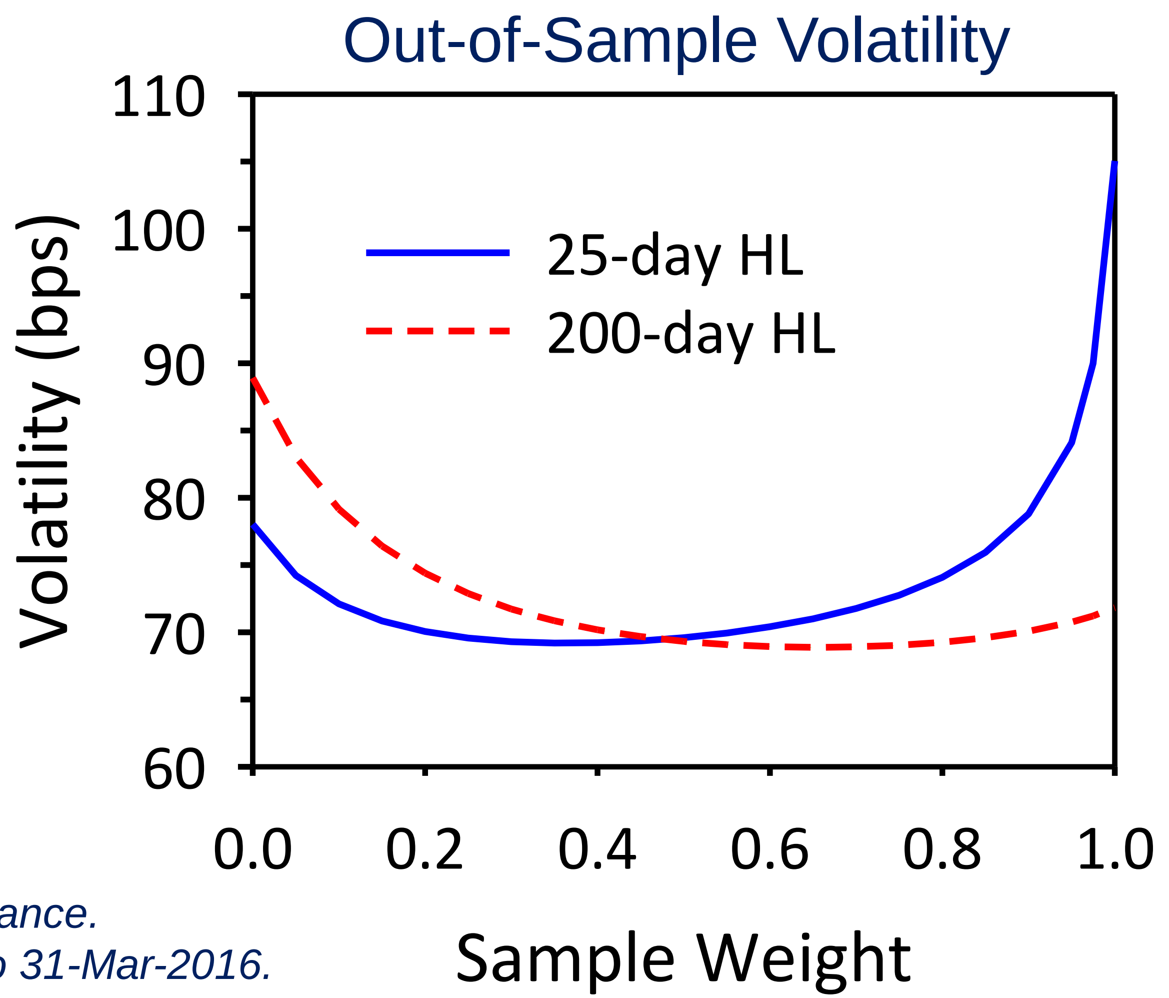
Ledoit and Wolf. *Improved Estimation of the Covariance matrix of Stock Returns,* Journal of Empirical Finance, December 2003, pp. 603-621

Menchero, Jose and Lei Ji. *Advances in Estimating Covariance Matrices,* Journal of Investment Management (to appear)

Bloomberg

Benefits of Blending (Stock Example)

- Form minimum-volatility fully invested portfolio of 100 stocks
- Compute sample correlation matrix (using 25d and 200d HL)
- Compute PCA model with one factor (25d and 200d HL)
- Blend the two covariance matrices
- Volatility is minimized at some intermediate blending weight
- Blended 25d HL model performs nearly as well as model with 200d HL
- Model with 25d HL assigns less weight to sample (0.4)
- Model with 200d HL assigns more weight to sample (0.7)

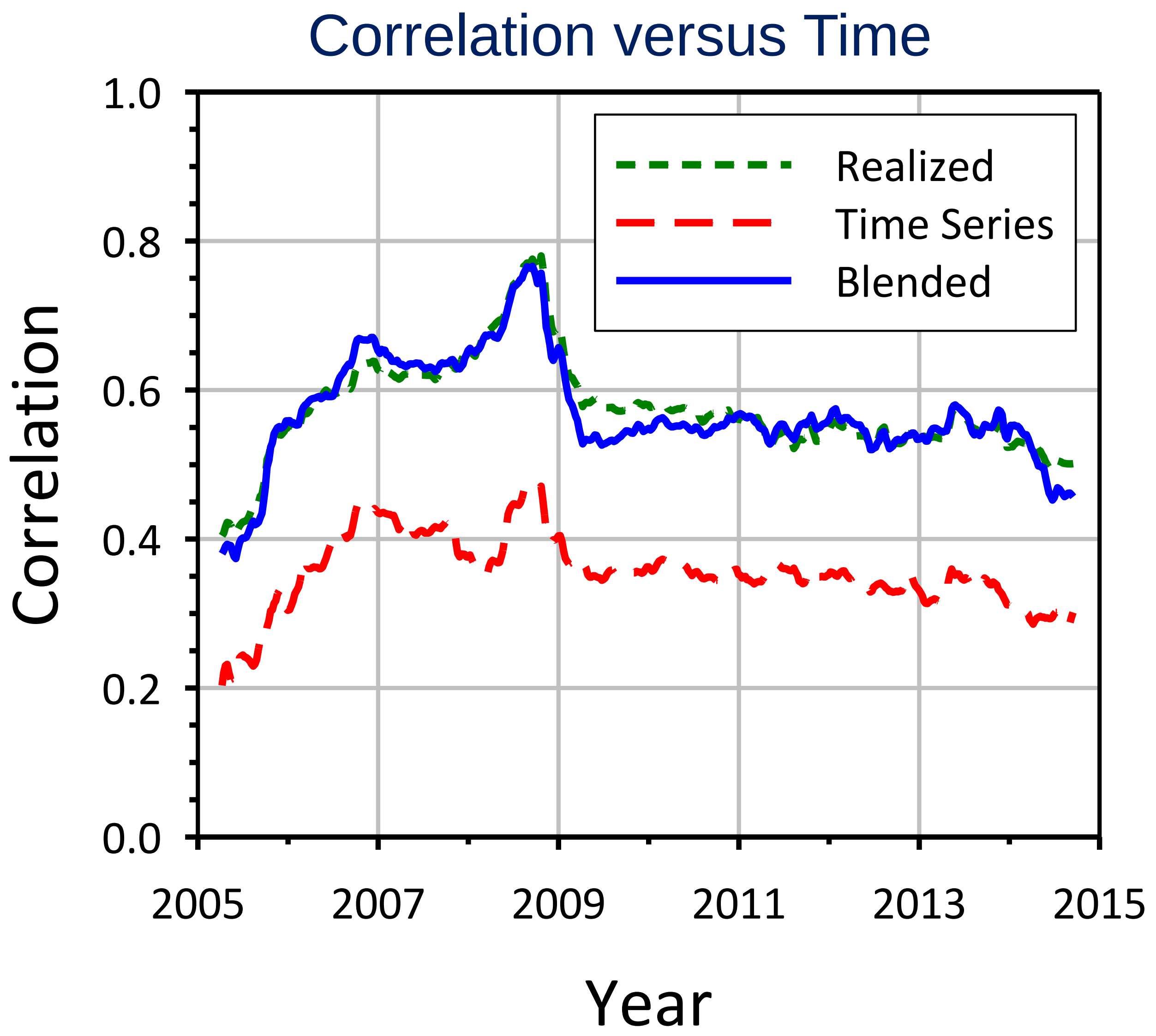


Top 100 US stocks. Daily rebalance.
Sample period: 27-Dec-2000 to 31-Mar-2016.

Bloomberg

Across Asset Classes (Equities vs Commodities)

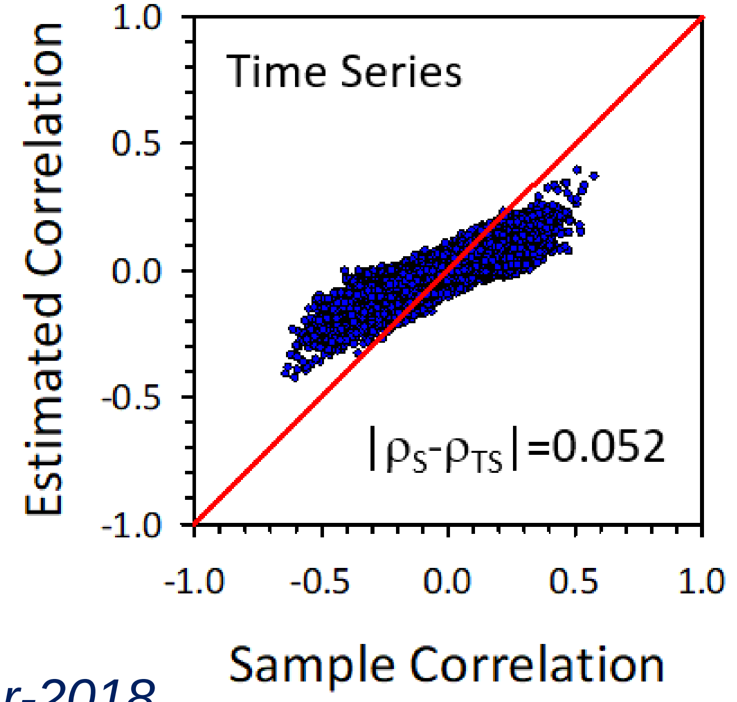
- Consider the correlation between the US energy factor (equity) and the crude-oil commodity factor (Brent shift)
- Plot predicted and realized correlations (52w
- Blended approach never deviates far from the realized (sample) correlation
- This makes the blended approach more accurate for making risk forecasts
- Time-series method systematically under-predicts correlation
- Result is typical and representative of other asset classes
- Suggests the time-series approach is incapable of fully capturing correlations

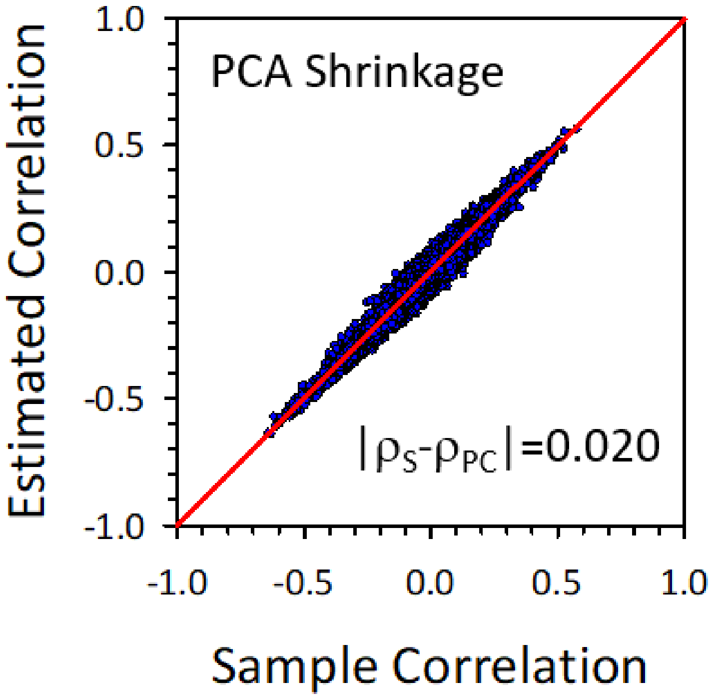


Bloomberg

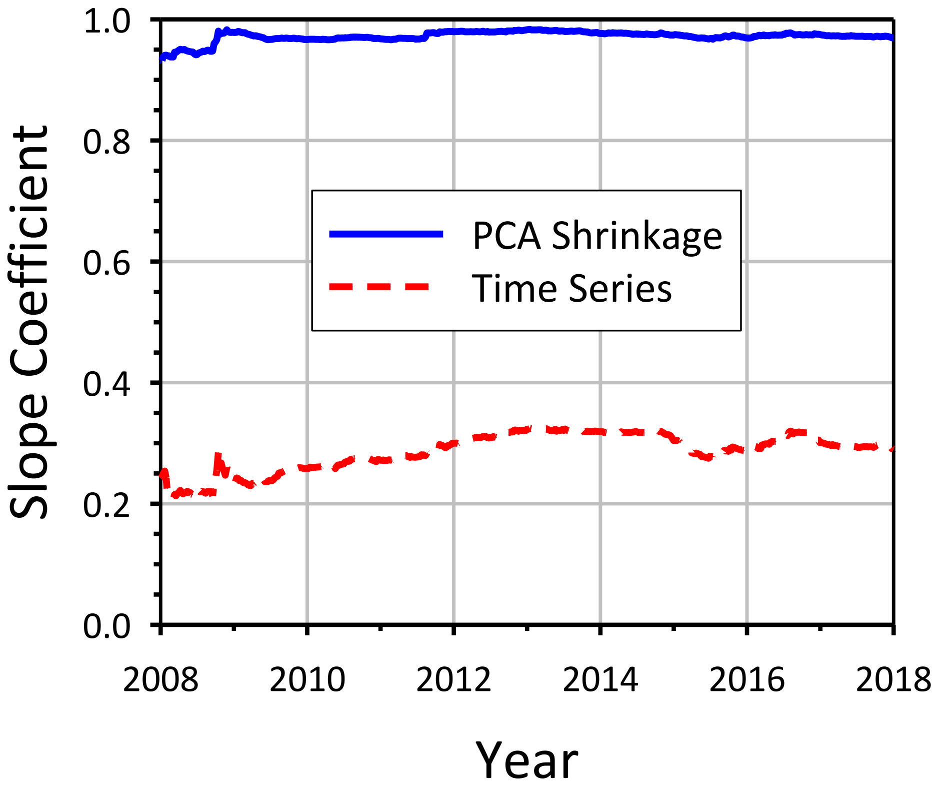
Correlation Scatterplots

- Goal: obtain well-conditioned correlation matrix (reliable for optimization) while deviating minimally from the sample correlation matrix (accurate risk forecasts)
- Conventional approach uses the "time-series method"
  - Tries to identify "global" factor returns to explain "local" correlations
- Time-series method tends to systematically underforecast ρ
  - Example: correlation between equity factors and fixed-income factors





Slope versus Time



Bloomberg

More Cross-Asset Results

- Report slope coefficients for correlation scatterplots across asset classes
- Includes 387 equity factors, 431 fixed income factors, and 222 commodity factors
- Time-series method
  - Large biases within each asset class
  - Biases are exacerbated across asset classes
  - Effective 71% shrinkage intensity between equities and fixed income
- Blended method
  - Near perfect slope coefficients within and across asset classes
  - Reliable correlation estimates for risk forecasting

	Time Series Method			PCA Shrinkage		
Asset Class	Equities	Fixed	Comm	Equities	Fixed	Comm
Equities	0.75	0.29	0.29	1.00	0.97	0.91
Fixed Income	0.29	0.62	0.21	0.97	1.01	0.94
Commodities	0.29	0.21	0.31	0.91	0.94	0.95

Full set of factors within each asset class

Empirical slope coefficients

Summary

- Sample Correlation
  - Great for making risk forecasts (except for optimized portfolios)
  - Not reliable for portfolio optimization
- Shrinking Correlations to Identity Matrix
  - Well-conditioned matrix (more robust for portfolio optimization)
  - Leads to biases in risk forecasts
- Blended Correlations
  - Blend sample with the PCA correlation matrix
  - Closely mimics sample correlation (good for risk forecasts)
  - Well-conditioned matrix (suitable for portfolio optimization)
- Time Series Method
  - Tends to systematically underforecast correlations
  - Not well suited for predicting portfolio volatility

Bloomberg