Covariance Matrix Estimation for Portfolio Selection: Markowitz Meets Goldilocks and Sharknadoes

Olivier Ledoit Michael Wolf

Department of Economics
University of Zurich

Disclaimer

This presentation was produced solely by Michael Wolf. The opinions and statements expressed herein are those of Michael Wolf are not necessarily the opinions of any other entity, including UBS AG and its affiliates. UBS AG and its affiliates accept no responsibility whatsoever for the accuracy, reliability or completeness of the information, statements or opinions contained in this presentation and will not be liable either directly or indirectly for any consequences, including any loss or damage, arising out of the use of or reliance on this presentation or any part thereof. Reproduced with permission.

The Oldest Problem in Finance

Problem Solved Already!

The Second-Oldest Problem in Finance

PORTFOLIO SELECTION*

HARRY MARKOWITZ
The Rand Corporation

THE PROCESS OF SELECTING a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio. This paper is concerned with the second stage. We first consider the rule that the investor does (or should) maximize discounted expected, or anticipated, returns. This rule is rejected both as a hypothesis to explain, and as a maximum to guide investment behavior. We next consider the rule that the investor does (or should) consider expected return a desirable thing *and* variance of return an undesirable thing. This rule has many sound points, both as a maxim for, and hypothesis about, investment behavior. We illustrate geometrically relations between beliefs and choice of portfolio according to the "expected returns—variance of returns" rule.

Outline

Outline

Markowitz: Theory vs. Practice

Markowitz (1952, JF) solved the problem of portfolio selection *in theory*.

His formulas use two inputs:

(i) the vector of means and

(ii) the covariance matrix

of the relevant asset returns.

In practice, these inputs are unknown and have to be estimated from data.

This problem has been

- a source of great frustration to portfolio managers
- a source of great (paper) creation to academic researchers

Markowitz: Early Days

Early practice:

- Estimate the two inputs by their sample counterparts
- These estimators are unbiased and MLEs under normality

Early critics:

- Jobson and Korkie (1980, JASA), Michaud (1989, FAJ), and Chopra and Ziemba (1993, JPM), among others, showed that this practice leads to unstable and underdiversified portfolios
- Therefore, such portfolios have poor out-of-sample performance

Michaud (1989, FAJ) coined the term "estimation error maximizers".

This is because Markowitz portfolios favor assets with

- large estimated means
- negative estimated covariances
- small estimated variances

Markowitz: Estimation Error Maximization

Vector of means:

- Financial returns are notoriously noisy
- Thus, sample means are very unreliable

Covariance matrix:

- Often the number of assets is comparable to the sample size
- In such a case, the sample covariance matrix is ill-conditioned
- This is a major reason for unstable portfolios, since the Markowitz formulas use the inverse of the covariance matrix

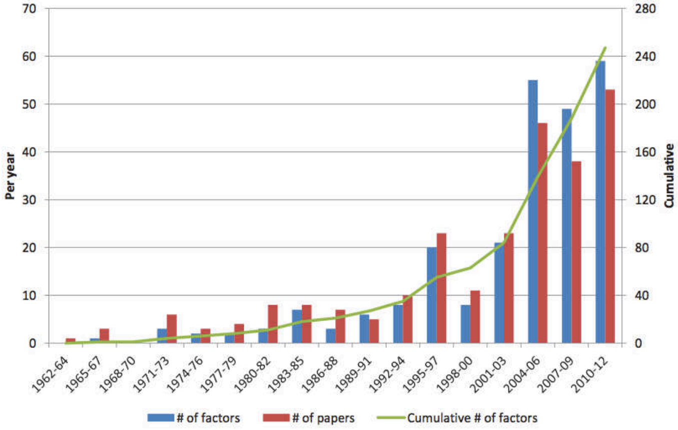
Example: 5 years of daily data on the Russell 1000

- Sample size: $T \approx 1,260$ days
- Dimension: $N = 1,000$ stocks

Vector of Means

Harvey, Liu, and Zhu (2016, RFS)

Figure 2: **Factors and Publications**



Outline

Line of Attack

Find a covariance matrix estimator that is optimal in a stylized setting of the Markowitz portfolio selection problem.

Key: Dimension of the space of candidate estimators

= Degrees of freedom

= Number of parameters to be estimated jointly

$\Longrightarrow$ Classification of the relevant literature.

Outline

Early Days: $O(N^2)$

Estimator: sample covariance matrix.

Number of free parameters: $N(N+1)/2$.

Seemed like a good idea:

- Unbiased estimator
- Maximum likelihood estimator (under normality)

Sad reality:

- Leads to unstable and underdiversified portfolios
- Such portfolios have poor out-of-sample properties

Unless $N \ll T$:

- Sample covariance matrix is ill-conditioned
- Too much estimation error

Fact: $O(N^2)$ is too big.

Extremist Reaction: $O(0)$

'Estimator': identity covariance matrix.

That is, do not estimate the covariance matrix at all!

Promoters:

- Fama and French (1993, JF): sort into deciles
- DeMiguel, Garlappi, and Uppal (2009, RFS) additionally abstain from the estimation of the vector of means: $1/N$ portfolio
- Brandt, Santa-Clara, and Valkanov (2009, RFS): portfolio spanned by vector(s) of means

Fact: $O(0)$ can be less bad than $O(N^2)$ but is too small.

Former State of the Art: $O(1)$

Synthesis of the first two approaches: linear shrinkage.

Estimator: convex combination of

- sample covariance matrix
- (multiple of the) identity matrix: shrinkage target

Proposed by Ledoit and Wolf (2004, JMVA).

Only one parameter:

- Shrinkage intensity (weight of the shrinkage target)

Linear shrinkage can be adapted to alternative shrinkage targets:

- Single-factor model, as in Ledoit and Wolf (2003, JEF)
- Constant-correlation model, as in Ledoit and Wolf (2004, JPM)

Fact: $O(1)$ is better than both $O(0)$ and $O(N^2)$.

Alternative $O(1)$ Methods

DeMiguel, Garlappi, Nogales, and Uppal (2009, MS):

- Norm-constrained portfolios
- Only for global mininum variance portfolio
- Needs cross validation
- Beats LW only in 1 out of 5 data sets

Frahm and Memmel (2010, JoE):

- Shrink portfolio weights to $1/N$
- Only for global mininum variance portfolio

Kan and Zhou (2007, JFQA):

- Weighted combination of sample portfolios with riskfree rate

Tu and Zhou (2011, JFE):

- Weighted combination of various portfolios with $1/N$

The last three proposals assume normality, do not work for $N > T$, and do not compare to LW.

The Big Picture

Tried so far:

- $O(0)$: passable
- $O(1)$: former state of the art
- $O(N^2)$: does not work, unless $N \ll T$

…anything missing?

Not Yet Tried: $O(N)$

Realization:

- Only (obvious) dimension that has not been tried yet
- Only chance to beat the former state-of-the-art $O(1)$
- But mathematically more challenging

Key insight:

- Optimal estimator in N-dimensional space should beat optimal estimator in 1- or 2-dimensional space (under nesting)
- Have to able to squash estimation error to obtain a consistent estimator

Goldilocks & the Three Bears

Goldilocks & Covariance Matrix Estimation

$O(0)$ and even $O(1)$ are too small due to misspecification error.

$O(N)$ is just right: largest number of free parameters that can be estimated consistently when $N \approx T$.

$O(N^2)$ is too big due to estimation error.

Goldilocks Payoff

Contributions relative to the $O(1)$ proposal of LW (2004, JMVA):

1. Search for optimal estimator in a much broader candidate space
2. Use an objective function that is tailor made for portfolio selection (instead of generic mean squared error)
3. Resulting portfolios have better out-of-sample properties (as demonstrated in Ledoit and Wolf (2017, RFS))

Outline

Starting Point

N denotes the number of assets and T denotes the sample size.

The sample covariance matrix S_T admits a spectral decomposition

$$S_T = U_T \Lambda_T U_T'$$

Here:

- U_T is an orthogonal matrix whose columns are the sample eigenvectors $(u_{T,1}, \ldots, u_{T,N})$
- Λ_T is a diagonal matrix whose diagonal entries are the sample eigenvalues $(\lambda_{T,1}, \ldots, \lambda_{T,N})$

Class of Estimators

Rotation Equivariance

- Observed $T \times N$ data matrix: Y_T
- W is an N-dimensional orthogonal / rotation matrix
- $\hat{\Sigma}_T := \hat{\Sigma}_T(Y_T)$ is a generic estimator of Σ_T
- It is rotation-equivariant if $\hat{\Sigma}_T(Y_T W) = W' \hat{\Sigma}_T(Y_T) W$

Without specific knowledge about Σ_T, rotation equivariance is a desirable property of an estimator.

We use the following class of rotation-equivariant estimators going back to Stein (1975, 1986).

$$\hat{\Sigma}_T := U_T D_T U_T' \quad \text{where} \quad D_T := \mathsf{Diag}(d_{T,1}, \ldots, d_{T,N}) \text{ is diagonal}$$

This is a class of dimension N.

Rotation-Equivariant Estimators

Generic estimator in the class $\hat{\Sigma}_T := U_T D_T U_T'$.

Keep the sample eigenvectors.

Shrink the N sample eigenvalues individually:

- $D_T := \mathsf{Diag}(d_T(\lambda_{T,1}), \ldots, d_T(\lambda_{T,N}))$
- Based on nonlinear shrinkage function $d_T : \mathbb{R} \to \mathbb{R}_+$

LW (2004, JMVA) only consider linear shrinkage function d_T.

Assumptions:

- The shrinkage function may be stochastic through dependence on the sample covariance matrix S_T
- It converges to a non-stochastic limiting shrinkage function d

Outline

Tailor-Made Loss Function

Mininum Variance Loss Function:

$$\mathcal{L}_{\mathrm{MV}}\left(\hat{\Sigma}_T, \Sigma_T\right) := \frac{\mathsf{Tr}\left(\hat{\Sigma}_T^{-1} \Sigma_T \hat{\Sigma}_T^{-1}\right) / N}{\left[\mathsf{Tr}\left(\hat{\Sigma}_T^{-1}\right) / N\right]^2} - \frac{1}{\mathsf{Tr}\left(\Sigma_T^{-1}\right) / N}$$

Roughly speaking, $\mathcal{L}_{\mathrm{MV}}$ represents the true variance of the portfolio with the minimum estimated variance, after suitable normalization.

Feasible Estimator

We use tools from random matrix theory and assume:

- $N/T \to c \in (0, \infty)$, as $T \to \infty$
- Data are independent and identically distributed (i.i.d.)
- Moment an distribution conditions
- Conditions on the eigenvalues of the true covariance matrix

Then:

- $\mathcal{L}_{\text{MV}}(\hat{\Sigma}_T, \Sigma_T)$ is non-stochastic in the limit
- Minimize the limiting expression with respect to d
- The optimal d, denoted by d^{o}, is an oracle
  (meaning it depends on unknown population quantities)
- Construct a consistent estimator of d^{o}, denoted by $\hat{d}_T^{\text{o}}$

Feasible Nonlinear Shrinkage Estimator:

$$S_T^{\text{o}} := U_T \hat{D}_T^{\text{o}} U_T' \quad \text{with} \quad \hat{D}_T^{\text{o}} := \mathsf{Diag}\big(\hat{d}_T^{\text{o}}(\lambda_{T,1}), \ldots, \hat{d}_T^{\text{o}}(\lambda_{T,N})\big)$$

Related Method: Eigenvalue Cleaning

A popular method used by practitioners is eigenvalue cleaning:

- Leave the large eigenvalues (“the signal”) unchanged
- Make all the small eigenvalues (“the noise”) equal to their average
- Usually based on the correlation matrix
- Also called eigenvalue denoising

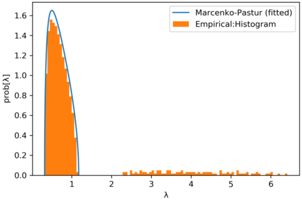
Remarks:

- This is a reasonable ad hoc method, but it is not optimal
- Large eigenvalues need to be adjusted too
  (though differently from linear shrinkage)
- Small eigenvalues should not be (exactly) equalized

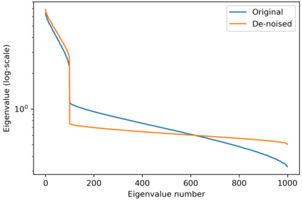
A typical unwarranted claim from the machine learning crowd:

Financial Correlations Are Extremely Noisy

- The econometric canon does not include methods to de-noise and de-tone correlation matrices
- As a result, most econometric studies reach spurious conclusions, supported by noise, not signal

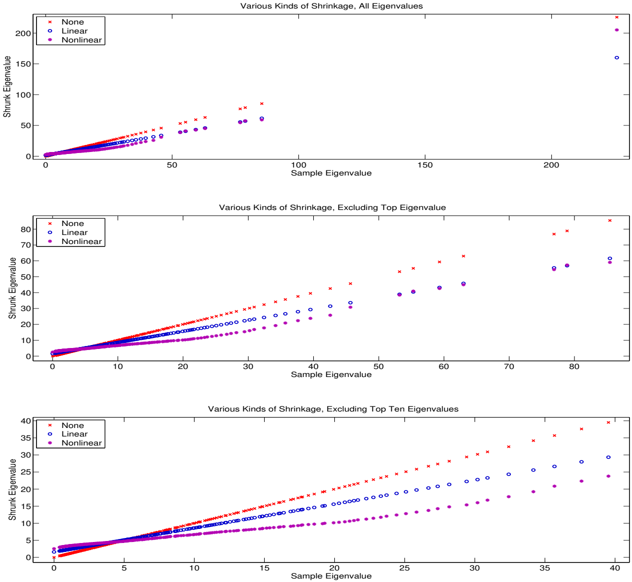


Almost all eigenvalues contained in a financial correlation matrix are associated with noise, not signal. Econometric studies estimate betas that reflect spurious relationships



Mathematical approaches can determine which eigenvalues must be treated numerically to prevent false discoveries, however those approaches are rarely used in econometric studies (N.B.: shrinkage fails to separate signal from noise)

Linear vs. Nonlinear Shrinkage ($N = 500$ stocks)



Outline

The Importance of Good Forecasts

Good forecasts of time-varying objects can make the difference between life and death.

Here is a weather-related example from the movie Sharknado 2:

The Importance of Good Forecasts

Good forecasts of time-varying objects can make the difference between life and death.

Here is a weather-related example from the movie Sharknado 2:

We now turn to forecasts of time-varying covariance matrices.

Motivation and Problem

Stylized fact:

- Asset returns often exhibit co-volatility clustering, at least at shorter frequencies, and are thus not i.i.d.

Common approach:

- Use a multivariate GARCH model to capture this effect

Problem:

- Such models suffer from the curse of dimensionality
- Applications are generally limited to $N \leq 100$ assets

DCC-NL Model

Univariate volatilities governed by a GARCH(1,1) process:

$$d_{i,t}^2 = \omega_i + a_i r_{i,t-1}^2 + b_i d_{i,t-1}^2$$

DCC model of Engle (2002, JBES) with correlation targeting:

$$Q_t = (1 - \alpha - \beta)\, C + \alpha\, s_{t-1} s'_{t-1} + \beta\, Q_{t-1} \tag{1}$$

where $s_{i,t} := r_{i,t}/d_{i,t}$, $s_t := (s_{1,t}, \ldots, s_{N,t})'$ and $C := \mathsf{Cov}(s_t)$.

Conditional correlation and covariance matrices then:

$$R_t := \mathsf{Diag}(Q_t)^{-1/2}\, Q_t\, \mathsf{Diag}(Q_t)^{-1/2}$$
$$H_t := D_t R_t D_t$$

with $r_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t)$.

Key: Use nonlinear shrinkage to estimate the targeting matrix C in (1).

$\Longrightarrow$ DCC-NL model of Engle, Ledoit, and Wolf (2019, JFEC)

Outline

Motivation & Problem

Stylized fact:

- Asset returns follow (more or less) a factor model
- Examples: CAPM, APT, and Fama-French factor models

Common approach:

- Use a structured estimator of the covariance matrix that is 'implied' by the assumed factor model

Problem:

- Which and how many factors to use?
- What if the factor model is misspecified?

Model and Implied Covariance Matrix

A factor model assumes that

$$r_{i,t} = \alpha_i + \beta_i' f_t + u_{i,t} \quad \text{with} \quad \mathbb{E}(u_{i,t}|f_t) = 0 \,,$$

where

- $f_t \in \mathbb{R}^K$ is a vector of factor returns
- α_i is an intercept and $\beta_i \in \mathbb{R}^K$ is a vector of factor loadings

Implied Covariance Matrix:

$$H_t = B'\Sigma_{f,t} B + \Sigma_{u,t} \quad \text{with} \quad B := [\beta_1, \ldots, \beta_N] \,.$$

A static version assumes $\Sigma_f \equiv \Sigma_{f,t}$ and $\Sigma_{u,t} \equiv \Sigma_u$, which implies $H_t \equiv H$.

Different Versions

In all versions:

- B is estimated by OLS, one asset at a time, yielding $\hat{B} := [\hat{\beta}_1, \ldots, \hat{\beta}_N]$ and residuals $\hat{u}_t := (\hat{u}_{1,t}, \ldots, \hat{u}_{N,t})'$
- $\hat{\Sigma}_f$ is the sample covariance matrix of the $\{f_t\}$ and $\hat{\Sigma}_{f,t} \equiv \hat{\Sigma}_f$
- Use $K = 1$ or $K = 5$ Fama-French factors

Exact Factor Model (EFM):

- Static model that assumes Σ_u is diagonal
- $\hat{\Sigma}_u$ is the diagonal part of the sample covariance matrix of the $\{\hat{u}_t\}$

Approximate Factor Model (AFM-NL):

- Static model that assumes nothing about Σ_u
- $\hat{\Sigma}_u$ is obtained by applying nonlinear shrinkage to the $\{\hat{u}_t\}$

Approximate Factor Model (AFM-DCC-NL):

- Dynamic model that assumes nothing about $\Sigma_{u,t}$
- $\hat{\Sigma}_{u,t}$ is obtained by applying DCC-NL to the $\{\hat{u}_t\}$

Averaged Forecasting

Problem:

- In our backtest analysis, we use daily data
- But we update the portfolios only once a month, that is, once every 21 trading days
- This creates a certain 'mismatch' for dynamic models

Solution of De Nard, Ledoit, and Wolf (2021, JFEC):

- Forecast the covariance matrix separately for all 21 trading days of the upcoming month
- Then average these 21 forecasts and use the averaged matrix for portfolio selection

To this end, we use standard proposals from the literature to forecast

(i) conditional volatilities based on GARCH(1,1) dynamics

(ii) conditional correlation matrices based on DCC-NL dynamics

Outline

Data & Portfolio Rules

Stocks:

- Download daily return data from CRSP
- Period: 01/01/1973–12/31/2017

Observed factors:

- Download return data for the five Fama-French factors
- Available on the website of Ken French

Updating:

- 21 consecutive trading days constitute one ‘month’
- Update portfolios on ‘monthly’ basis

Out-of-sample period:

- Start out-of-sample investing on 01/16/1978
- This results in 10,080 daily returns (over 480 ‘months’)

Data & Portfolio Rules

Portfolio sizes:

- We consider $N \in \{100, 500, 1000\}$

Portfolio constituents:

- Select new constituents at the beginning of each month
- If there are pairs of highly correlated stocks ($r > 0.95$), kick out the stock with lower market capitalization
- Find the N largest remaining stocks that have
  - (i) a nearly complete 1260-day return history
  - (ii) a complete 21-day return future

Estimation:

- Use previous $T = 1260$ days to estimate the covariance matrix

Performance Measures

All measures are based on the 10,080 out-of-sample returns and are annualized for convenience.

Performance measures:

- AV: Average
- SD: Standard deviation
- IR: Information ratio

Outline

Problem & Solutions

Problem Formulation:

$$\min_w w' H_t w$$
$$\text{subject to} \quad w'\mathbb{1} = 1$$

(where $\mathbb{1}$ is a conformable vector of ones)

Analytical Solution:

$$w^* = \frac{H_t^{-1}\mathbb{1}}{\mathbb{1}' H_t^{-1}\mathbb{1}}$$

Feasible Solution:

$$\hat{w} := \frac{\hat{H}_t^{-1}\mathbb{1}}{\mathbb{1}' \hat{H}_t^{-1}\mathbb{1}}$$

Performance Measures

	$N = 100$			$N = 500$			$N = 1000$		
	AV	SD	IR	AV	SD	IR	AV	SD	IR
				Structure-Free Models					
$1/N$	12.82	17.40	0.74	13.86	16.83	0.82	14.36	16.85	0.85
NL	11.94	11.74	1.02	11.91	8.63	1.38	12.28	7.45	1.65
DCC-NL	11.62	11.59	1.00	12.57	8.26	1.52	12.84	6.93	1.85
				Exact Factor Models					
EFM1	13.06	14.12	0.93	12.52	12.14	1.03	13.35	10.97	1.22
EFM5	13.02	12.68	1.03	12.68	10.97	1.16	12.90	9.72	1.33
				Approximate Factor Models					
POET	12.04	11.98	1.00	11.86	8.48	1.40	13.09	7.82	1.67
AFM1-NL	11.97	11.75	1.02	11.90	8.63	1.38	12.28	7.45	1.65
AFM5-NL	11.95	11.76	1.02	11.88	8.63	1.38	12.20	7.45	1.64
AFM1-DCC-NL	11.55	**11.56**	1.00	12.65	**8.11**	1.56	13.31	**6.61**	2.01
AFM5-DCC-NL	11.53	11.64	0.99	12.53	8.18	1.53	12.92	6.65	1.94

Note: In the columns labeled "SD", the best numbers are in **blue**.

Performance Measures ($N = 1000$)

	AV	SD	IR
	Structure-Free Models		
$1/N$	14.36	16.85	0.85
NL	12.28	7.45	1.65
DCC-NL	12.84	6.93	1.85
	Exact Factor Models		
EFM1	13.35	10.97	1.22
EFM5	12.90	9.72	1.33
	Approximate Factor Models		
POET	13.09	7.82	1.67
AFM1-NL	12.28	7.45	1.65
AFM5-NL	12.20	7.45	1.64
AFM1-DCC-NL	13.31	**6.61**	2.01
AFM5-DCC-NL	12.92	6.65	1.94

Outline

Problem & Solutions

Problem Formulation:

$$\min_w w' H_t w$$
$$\text{subject to} \quad w' m_t = b \quad \text{and}$$
$$w' \mathbb{1} = 1$$

(where m_t is a signal and b is a target expected return)

Analytical Solution:

$$w^* = c_1 H_t^{-1} \mathbb{1} + c_2 H_t^{-1} m$$
$$\text{where} \quad c_1 := \frac{C - bB}{AC - B^2} \quad \text{and} \quad c_2 := \frac{bA - B}{AC - B^2}$$
$$\text{with} \quad A := \mathbb{1}' H_t^{-1} \mathbb{1} \quad B := \mathbb{1}' H_t^{-1} b \quad \text{and} \quad C := m' H_t^{-1} m$$

Feasible solution $\widehat{w}$ replaces H_t with an estimator $\widehat{H}_t$.

Signal and Target Expected Return

For the signal we use momentum:

- Return over the last 12 months, excluding the most recent month
- Has been around for a long time and is non-controversial
- Also can be computed from observed return data alone, whereas most other signals need outside information

Simple-minded benchmark:

- Equally invest in the top 20% of the stocks
- Called EW-TQ for "equally-weighted top-quintile"
- In the spirit of portfolio sorts à la Fama and French

Target expected return:

- We take b to be the expected return of the EW-TQ portfolio according to momentum

Performance Measures

	$N = 100$			$N = 500$			$N = 1000$		
	AV	SD	IR	AV	SD	IR	AV	SD	IR
Structure-Free Models									
EW-TQ	16.55	21.33	0.78	16.85	20.24	0.83	17.55	20.30	0.87
NL	14.76	14.16	1.04	14.54	10.10	1.44	15.00	8.75	1.71
DCC-NL	14.95	14.13	1.06	14.87	9.51	1.56	14.82	7.95	1.86
Exact Factor Models									
EFM1	15.37	16.50	0.93	15.52	13.93	1.11	16.33	12.78	1.28
EFM5	15.22	15.49	0.98	15.76	12.80	1.23	15.94	11.39	1.40
Approximate Factor Models									
POET	14.53	14.33	1.01	14.28	10.02	1.43	15.45	9.10	1.70
AFM1-NL	14.79	14.16	1.04	14.52	10.09	1.38	15.00	8.75	1.72
AFM5-NL	14.78	14.17	1.04	14.48	10.10	1.44	14.90	8.75	1.70
AFM1-DCC-NL	14.69	14.02	1.05	15.24	9.46	**1.61**	15.76	7.84	**2.01**
AFM5-DCC-NL	14.58	14.09	1.04	14.97	9.58	1.56	15.28	7.91	1.93

Note: In the columns labeled “IR”, the best numbers are in **blue**.

Academia vs. Real Life

Our ‘simple-minded’ back-tests are meant to identify the best covariance matrix estimator, not to evaluate realistic trading strategies.

Real-life portfolio managers face many additional constraints concerning gross-exposure, factor exposure, trading costs, etc.

But they still benefit from using the best covariance matrix estimator.

Back-Tests vs. Monte Carlo Studies

Some consider Monte Carlo studies more informative than back-tests.

Often it is said that "history will not repeat itself", but then the DGP of the Monte Carlo study is calibrated based on historical inputs ...

Monte Carlo studies offer flexibility to make one's own methods look good compared to other methods.

Outline

Conclusion

Naïve benchmarks based on sorting and equal-weighting
can be outperformed easily, at least when investing in individual stocks.

Dynamic covariance matrix estimators outperform static ones.

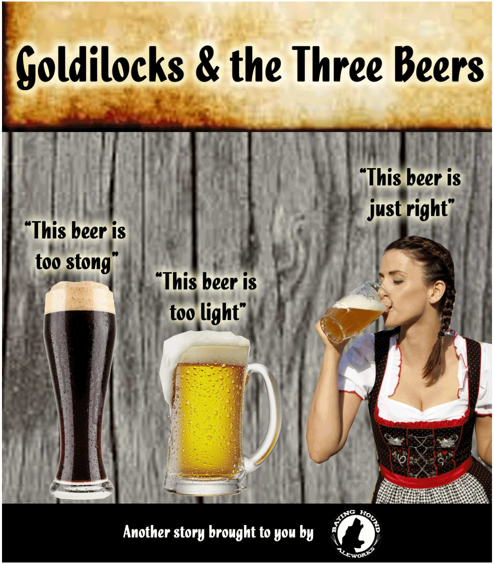
Injecting factor structure pays off, but no need to go beyond
the market factor if the residual covariance matrix is handled smartly.

This is good news for investors outside of the US.

The overall winner is AFM1-DCC-NL:

- Uses only the market factor
- Models the residual covariance matrix with DCC-NL

The Goldilocks Principle Has Universal Appeal!



De Nard, G., Ledoit, O., and Wolf, M. (2021). Factor models for portfolio selection in large dimensions: The good, the better and the ugly. *Journal of Financial Econometrics*. Forthcoming.

Engle, R. F. (2002). Dynamic conditional correlation — a simple class of multivariate GARCH models. *Journal of Business & Economics Statistics*, 20:339–350.

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large dynamic covariance matrices. *Journal of Business & Economic Statistics*, 37(2):363–375.

Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10(5):603–621.

Ledoit, O. and Wolf, M. (2004a). Honey, I shrunk the sample covariance matrix. *Journal of Portfolio Management*, 30(4):110–119.

Ledoit, O. and Wolf, M. (2004b). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.

Ledoit, O. and Wolf, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Annals of Statistics*, 40(2):1024–1060.

Ledoit, O. and Wolf, M. (2015). Spectrum estimation: A unified framework for covariance matrix estimation and PCA in large dimensions. *Journal of Multivariate Analysis*, 139(2):360–384.

Ledoit, O. and Wolf, M. (2017). Nonlinear shrinkage of the covariance matrix for portfolio selection: Markowitz meets Goldilocks. *Review of Financial Studies*, 30(12):4349–4388.

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