University of Zurich[UZH]

ETH zürich

Combining the MGHyp Distribution with Nonlinear Shrinkage in Modeling Financial Asset Returns

Simon Hediger, Jeffrey Näf

University of Zurich, ETH Zurich

04.06.2022

1. Methodology
2. Simulation
3. Portfolio Application
4. References

On the one hand

- Flexible elliptical mixture distributions:

$$\mathbf{Y} = \boldsymbol{\mu} + G^{1/2}\mathbf{H}^{1/2}\boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I})$ and G is a random variable with support $(0, +\infty)$, independent of $\boldsymbol{\epsilon}$.

[1] The MVG is a special case of the COMFORT model in Paolella and Polak (2015).

On the one hand

- Flexible elliptical mixture distributions:

$$\mathbf{Y} = \boldsymbol{\mu} + G^{1/2}\mathbf{H}^{1/2}\boldsymbol{\epsilon}, \tag{1}$$

  where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I})$ and G is a random variable with support $(0, +\infty)$, independent of $\boldsymbol{\epsilon}$.
- For example in the paper we focus specifically on the Multivariate Variance-Gamma (MVG), where[1]

$$G \sim \text{Gamma}(\lambda, 1).$$

[1] The MVG is a special case of the COMFORT model in Paolella and Polak (2015).

On the one hand

- Flexible elliptical mixture distributions:

$$\mathbf{Y} = \boldsymbol{\mu} + G^{1/2}\mathbf{H}^{1/2}\boldsymbol{\epsilon}, \tag{1}$$

  where $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \mathbf{I})$ and G is a random variable with support $(0, +\infty)$, independent of $\boldsymbol{\epsilon}$.

- For example in the paper we focus specifically on the Multivariate Variance-Gamma (MVG), where[1]

$$G \sim \text{Gamma}(\lambda, 1).$$

- In general one might write $\mathbf{Y} \sim GM(\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$, where $\boldsymbol{\theta}_L$ collects all parameters determining the distribution of G.

[1] The MVG is a special case of the COMFORT model in Paolella and Polak (2015).

On the one hand

- Crucially for some of these distributions, such as multivariate t and MVG, efficient EM algorithms are available for parameter estimation.

On the one hand

- Crucially for some of these distributions, such as multivariate t and MVG, efficient EM algorithms are available for parameter estimation.
- The algorithms impute the latent variable G by its conditional expectation and then use Gaussian estimates for $\boldsymbol{\mu}$ and $\mathbf{H}$.

On the one hand

- Crucially for some of these distributions, such as multivariate t and MVG, efficient EM algorithms are available for parameter estimation.
- The algorithms impute the latent variable G by its conditional expectation and then use Gaussian estimates for $\boldsymbol{\mu}$ and $\mathbf{H}$.
- Let in the following
  1. $\boldsymbol{\theta} = (\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$,
  2. T the number of observations,
  3. N the number of dimensions (assets).

General ECME algorithm

E-step: For $t = 1, \ldots, T$, calculate $\hat{G}_t^{-1} = \mathbb{E}\left[G_t^{-1} \mid \mathbf{Y}_t, \hat{\theta}\right]$.

CM1-step: Update $\boldsymbol{\mu}, \mathbf{H}$ by first obtaining the weighted mean

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{t=1}^{T} \hat{G}_t^{-1/2} \mathbf{y}_t}{\sum_{t=1}^{T} \hat{G}_t^{-1/2}}. \tag{2}$$

Then, with $\hat{\epsilon}_t = \hat{G}_t^{-1/2} (\mathbf{y}_t - \hat{\boldsymbol{\mu}})$, calculate

$$\hat{\mathbf{H}} = \frac{1}{T} \sum_{i=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t^{\top}. \tag{3}$$

CM2-step: Given the CM1-step updates of $\boldsymbol{\mu}, \mathbf{H}$, obtain new updates of $\boldsymbol{\theta}_L$ by numerically maximizing the log-likelihood function $\ln L_{\mathbf{Y}}(\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$ with respect to $\boldsymbol{\theta}_L$.

On the other hand

- The above algorithm monotonically increases the likelihood $L_{\mathbf{Y}}$ in each step.

On the other hand

- The above algorithm monotonically increases the likelihood $L_{\mathbf{Y}}$ in each step.
- The key is that both the CM1- and CM2-step *optimize a likelihood*:
  - The CM1-Step maximizes a Gaussian likelihood weighted with the imputed $\hat{G}_t$,
  - The CM2-Step maximizes the marginal likelihood.

On the other hand

- The above algorithm monotonically increases the likelihood $L_{\mathbf{Y}}$ in each step.
- The key is that both the CM1- and CM2-step *optimize a likelihood*:
  - The CM1-Step maximizes a Gaussian likelihood weighted with the imputed $\hat{G}_t$,
  - The CM2-Step maximizes the marginal likelihood.
- This idea should guide the construction of any new EM algorithm.

On the other hand

- As usual, as N approaches T the estimation of the dispersion matrix in (3) becomes problematic.

On the other hand

- As usual, as N approaches T the estimation of the dispersion matrix in (3) becomes problematic.
- Many shrinkage methods exist: Linear shrinkage, Factor modelling.

On the other hand

- As usual, as N approaches T the estimation of the dispersion matrix in (3) becomes problematic.
- Many shrinkage methods exist: Linear shrinkage, Factor modelling.
- Nonlinear Shrinkage (NL) of the covariance matrix was introduced and subsequently (computationally) refined in Ledoit and Wolf (2012, 2020a,b).

Nonlinear Shrinkage

- NL is extremely powerful, both in theory and in practice, and especially in finance applications.

Nonlinear Shrinkage

- NL is extremely powerful, both in theory and in practice, and especially in finance applications.
- Consider the eigenvalue decomposition of the sample covariance matrix $\hat{\mathbf{\Sigma}}$:

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^\top. \quad (4)$$

Nonlinear Shrinkage

- NL is extremely powerful, both in theory and in practice, and especially in finance applications.
- Consider the eigenvalue decomposition of the sample covariance matrix $\hat{\mathbf{\Sigma}}$:

$$\hat{\mathbf{\Sigma}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^\top. \tag{4}$$

- Both linear and nonlinear shrinkage shrink the *eigenvalues* of the sample covariance matrix $\hat{\mathbf{\Sigma}}$.

Nonlinear Shrinkage

- NL is extremely powerful, both in theory and in practice, and especially in finance applications.
- Consider the eigenvalue decomposition of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$:

$$\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top. \tag{4}$$

- Both linear and nonlinear shrinkage shrink the *eigenvalues* of the sample covariance matrix $\hat{\boldsymbol{\Sigma}}$.
- NL shrinkage (asymptotically) solves the problem

$$\boldsymbol{\Lambda}^* = \underset{\boldsymbol{\Lambda}\text{diagonal}}{\arg\min} \operatorname{Tr}\left[(\boldsymbol{\Sigma} - \hat{\mathbf{U}}\boldsymbol{\Lambda}\hat{\mathbf{U}}^\top)^2\right]$$

Nonlinear Shrinkage

- "Nonlinear" shrinkage, because the eigenvalues get shrunken in a nonlinear way.

Nonlinear Shrinkage

- "Nonlinear" shrinkage, because the eigenvalues get shrunken in a nonlinear way.
- This is in contrast to linear shrinkage, where

$$\boldsymbol{\Lambda}^* = (1-\rho)\hat{\boldsymbol{\Lambda}} + \rho \mathbf{I}. \tag{5}$$

Nonlinear Shrinkage

- "Nonlinear" shrinkage, because the eigenvalues get shrunken in a nonlinear way.
- This is in contrast to linear shrinkage, where

$$\mathbf{\Lambda}^* = (1-\rho)\hat{\mathbf{\Lambda}} + \rho\mathbf{I}. \tag{5}$$

- Expressions are based on random matrix theory and are quite complicated.

Nonlinear Shrinkage into the EM

- Nonlinear shrinkage is quite complicated.

Nonlinear Shrinkage into the EM

- Nonlinear shrinkage is quite complicated.
- It appears unclear how to write it as a likelihood problem, especially in finite sample.

Nonlinear Shrinkage into the EM

- Nonlinear shrinkage is quite complicated.
- It appears unclear how to write it as a likelihood problem, especially in finite sample.
- Main Idea: Shrink the eigenvalues beforehand and use EM given a fixed $\mathbf{\Lambda}^*$.

General EM algorithm

E-step: For $t = 1, \ldots, T$, calculate $\hat{G}_t^{-1} = \mathbb{E}\left[G_t^{-1} \mid \mathbf{Y}_t, \hat{\theta}\right]$.

CM1-step: Update $\boldsymbol{\mu}, \mathbf{H}$ by first obtaining the weighted mean

$$\hat{\boldsymbol{\mu}} = \frac{\sum_{t=1}^{T} \hat{G}_t^{-1/2} \mathbf{y}_t}{\sum_{t=1}^{T} \hat{G}_t^{-1/2}}. \tag{2}$$

Then, with $\hat{\epsilon}_t = \hat{G}_t^{-1/2} (\mathbf{y}_t - \hat{\boldsymbol{\mu}})$, calculate

$$\hat{\mathbf{H}} = \frac{1}{T} \sum_{i=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t^{\top}. \tag{3}$$

CM2-step: Given the CM1-step updates of $\boldsymbol{\mu}, \mathbf{H}$, obtain new updates of $\boldsymbol{\theta}_L$ by numerically maximizing the log-likelihood function $\ln L_{\mathbf{Y}}(\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$ with respect to $\boldsymbol{\theta}_L$.

General idea

- We always assume $\boldsymbol{\Lambda}^*$ has ordered elements $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_N$.

General idea

- We always assume $\mathbf{\Lambda}^*$ has ordered elements $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_N$.
- Now the eigenvalues are fixed, the estimator $\mathbf{H}$ is of the form

$$\mathbf{H} = \mathbf{V}\mathbf{\Lambda}^*\mathbf{V}^\top.$$

General idea

- We always assume $\mathbf{\Lambda}^*$ has ordered elements $\delta_1 \geq \delta_2 \geq \ldots \geq \delta_N$.
- Now the eigenvalues are fixed, the estimator $\mathbf{H}$ is of the form

$$\mathbf{H} = \mathbf{V}\mathbf{\Lambda}^*\mathbf{V}^\top.$$

- Thus the estimation is over orthogonal matrices $\mathbf{V}$: $\mathbf{V} : \mathbf{V}^\top\mathbf{V} = \mathbf{V}\mathbf{V}^\top = \mathbf{I}$.

Nonlinear Shrinkage into the EM

- Main Idea: Shrink the eigenvalues beforehand and use EM given fixed $\mathbf{\Lambda}^*$.

Nonlinear Shrinkage into the EM

- Main Idea: Shrink the eigenvalues beforehand and use EM given fixed $\boldsymbol{\Lambda}^*$.
- Problem 1: How to get a monotone algorithm given a fixed $\boldsymbol{\Lambda}^*$?

Nonlinear Shrinkage into the EM

- Main Idea: Shrink the eigenvalues beforehand and use EM given fixed $\mathbf{\Lambda}^*$.
- Problem 1: How to get a monotone algorithm given a fixed $\mathbf{\Lambda}^*$?
- Problem 2: How to obtain $\mathbf{\Lambda}^*$?

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\mathbf{\Lambda}^*$?

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\boldsymbol{\Lambda}^*$?
- Need to optimize the likelihood

$$\ln L_{\mathbf{Y}|G}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left\{K\ln(2\pi G_t) + \ln(|\boldsymbol{\Lambda}^*|) + \hat{\epsilon}_t^\top \mathbf{V}(\boldsymbol{\Lambda}^*)^{-1}\mathbf{V}^\top \hat{\epsilon}_t\right\}$$

over $\mathbf{V}$.

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\boldsymbol{\Lambda}^*$?
- Need to optimize the likelihood

$$\ln L_{\mathbf{Y}|G}(\theta) = -\frac{1}{2}\sum_{t=1}^{T}\left\{K\ln(2\pi G_t) + \ln(|\boldsymbol{\Lambda}^*|) + \hat{\epsilon}_t^\top \mathbf{V}(\boldsymbol{\Lambda}^*)^{-1}\mathbf{V}^\top \hat{\epsilon}_t\right\}$$

over $\mathbf{V}$.

Proposition

$$\underset{\mathbf{V}:\mathbf{V}^\top\mathbf{V}=\mathbf{V}\mathbf{V}^\top=\mathbf{I}}{\arg\max} \ln L_{\mathbf{Y}|G}(\mu, \mathbf{H}) = \hat{\mathbf{U}}, \tag{6}$$

where $\hat{\mathbf{U}}$ *is part of the eigenvalue decomposition of* $\hat{\mathbf{H}}$:

$$\hat{\mathbf{H}} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top.$$

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\mathbf{\Lambda}^*$?
- We still maximize a (constrained) Gaussian likelihood, when choosing

$$\mathbf{V} = \hat{\mathbf{U}}$$

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\mathbf{\Lambda}^*$?
- We still maximize a (constrained) Gaussian likelihood, when choosing

$$\mathbf{V} = \hat{\mathbf{U}}$$

- The obtained estimator is then

$$\tilde{\mathbf{H}} = \hat{\mathbf{U}}\mathbf{\Lambda}^*\hat{\mathbf{U}}^\top \tag{7}$$

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\mathbf{\Lambda}^*$?
- We still maximize a (constrained) Gaussian likelihood, when choosing

$$\mathbf{V} = \hat{\mathbf{U}}$$

- The obtained estimator is then

$$\tilde{\mathbf{H}} = \hat{\mathbf{U}} \mathbf{\Lambda}^* \hat{\mathbf{U}}^\top \tag{7}$$

$\Rightarrow$ This is exactly the NL shrinkage approach, but now on the "filtered" values, that should ideally be more Gaussian.

Problem 1

- Problem 1: How to get a monotone algorithm given a fixed $\mathbf{\Lambda}^*$?
- We still maximize a (constrained) Gaussian likelihood, when choosing

$$\mathbf{V} = \hat{\mathbf{U}}$$

- The obtained estimator is then

$$\tilde{\mathbf{H}} = \hat{\mathbf{U}} \mathbf{\Lambda}^* \hat{\mathbf{U}}^\top \tag{7}$$

$\Rightarrow$ This is exactly the NL shrinkage approach, but now on the "filtered" values, that should ideally be more Gaussian.

$\Rightarrow$ If we run the above EM algorithm exchanging (3) with (7), the resulting algorithm monotonically increases the likelihood $L_{\mathbf{Y}}$ in each step.

Problem 2

- Problem 2: how to obtain $\mathbf{\Lambda}^*$?

Problem 2

- Problem 2: how to obtain $\mathbf{\Lambda}^*$?
- For our approach to make sense, $\mathbf{\Lambda}^*$ should be the shrunken eigenvalues of $\mathbf{H}$ in

$$\mathbf{Y} = \boldsymbol{\mu} + G^{1/2}\mathbf{H}^{1/2}\epsilon.$$

Problem 2

- Problem 2: how to obtain $\mathbf{\Lambda}^*$?
- For our approach to make sense, $\mathbf{\Lambda}^*$ should be the shrunken eigenvalues of $\mathbf{H}$ in

$$\mathbf{Y} = \boldsymbol{\mu} + G^{1/2}\mathbf{H}^{1/2}\epsilon.$$

- However: before having an estimation of $\mathbb{E}[G]$, only an estimate of

$$\text{Cov}(\mathbf{Y}) = \mathbf{\Sigma} = \mathbb{E}[G] \cdot \mathbf{H},$$

is available.

Problem 2

- Idea: simply standardize the data by the estimated variances.

Problem 2

- Idea: simply standardize the data by the estimated variances.
- Let

$$\boldsymbol{\sigma} = \begin{pmatrix} \sqrt{\mathrm{Var}(Y_1)} & 0 & \cdots & 0 \\ 0 & \sqrt{\mathrm{Var}(Y_2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\mathrm{Var}(Y_N)} \end{pmatrix},$$

$$\mathbf{S} = \begin{pmatrix} \sqrt{H_{11}} & 0 & \cdots & 0 \\ 0 & \sqrt{H_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{H_{NN}} \end{pmatrix}$$

Problem 2

- Idea: simply standardize the data by the estimated variances.

Problem 2

- Idea: simply standardize the data by the estimated variances.
- We consider the decomposition:

$$\mathbf{H} = \mathbf{S}\boldsymbol{\Gamma}\mathbf{S}, \tag{8}$$

where $\boldsymbol{\Gamma}$ is the correlation matrix of $\mathbf{H}^{1/2}\epsilon$.

Problem 2

- Idea: simply standardize the data by the estimated variances.
- We consider the decomposition:

$$\mathbf{H} = \mathbf{S}\boldsymbol{\Gamma}\mathbf{S}, \tag{8}$$

where $\boldsymbol{\Gamma}$ is the correlation matrix of $\mathbf{H}^{1/2}\boldsymbol{\epsilon}$.

- Notice that $\boldsymbol{\sigma} = \mathbf{S}\mathbb{E}[G]^{1/2}$ and so

$$\text{Corr}(\mathbf{Y}_t) = \boldsymbol{\sigma}^{-1}\mathbb{E}[G]\mathbf{H}_t\boldsymbol{\sigma}^{-1} = \boldsymbol{\Gamma}.$$

Problem 2

- Notice that $\boldsymbol{\sigma} = \mathbf{S}\mathbb{E}[G]^{1/2}$ and so

$$\mathrm{Corr}(\mathbf{Y}_t) = \boldsymbol{\sigma}^{-1}\mathbb{E}[G]\mathbf{H}_t\boldsymbol{\sigma}^{-1} = \boldsymbol{\Gamma}.$$

Problem 2

- Notice that $\boldsymbol{\sigma} = \mathbf{S}\mathbb{E}[G]^{1/2}$ and so

$$\mathrm{Corr}(\mathbf{Y}_t) = \boldsymbol{\sigma}^{-1}\mathbb{E}[G]\mathbf{H}_t\boldsymbol{\sigma}^{-1} = \boldsymbol{\Gamma}.$$

- Utilizing this intuition, we apply NL shrinkage on the estimated correlation matrix of $\mathbf{Y}$ to obtain $\boldsymbol{\Lambda}^*$.

Combining the ideas: Step 0

- For $\mathbf{Y} \sim GM(\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$ and $\mathbf{X} = \boldsymbol{\sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$,

$$\mathbf{X} \sim GM(\mathbf{0}, \boldsymbol{\Gamma}/\mathbb{E}[G], \boldsymbol{\theta}_L).$$

Combining the ideas: Step 0

- For $\mathbf{Y} \sim GM(\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$ and $\mathbf{X} = \boldsymbol{\sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$,

$$\mathbf{X} \sim GM(\mathbf{0}, \boldsymbol{\Gamma}/\mathbb{E}[G], \boldsymbol{\theta}_L).$$

- Thus we first standardize the data with the estimated standard deviations $\hat{\boldsymbol{\sigma}}$ and mean $\bar{\mathbf{y}}$:

$$\mathbf{x}_t = \hat{\boldsymbol{\sigma}}^{-1}(\mathbf{y}_t - \bar{\mathbf{y}}). \tag{9}$$

Combining the ideas: Step 0

- For $\mathbf{Y} \sim GM(\boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\theta}_L)$ and $\mathbf{X} = \boldsymbol{\sigma}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$,

$$\mathbf{X} \sim GM(\mathbf{0}, \boldsymbol{\Gamma}/\mathbb{E}[G], \boldsymbol{\theta}_L).$$

- Thus we first standardize the data with the estimated standard deviations $\hat{\boldsymbol{\sigma}}$ and mean $\bar{\mathbf{y}}$:

$$\mathbf{x}_t = \hat{\boldsymbol{\sigma}}^{-1}(\mathbf{y}_t - \bar{\mathbf{y}}). \tag{9}$$

- As mentioned we then obtain $\boldsymbol{\Lambda}^*$ from the covariance estimator of the sample $\mathbb{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$.

Combining the ideas: ECME

E-step: For $t = 1, \dots, T$, calculate $\hat{G}_t^{-1} = \mathbb{E}\left[G_t^{-1} \mid \mathbf{Y}_t, \hat{\theta}\right]$.

CM1-step: Update $\hat{\mu}_X$, $\tilde{\Gamma}$ by first obtaining the weighted mean

$$\hat{\mu}_X = \frac{\sum_{t=1}^{T} \hat{G}_t^{-1/2}\mathbf{x}_t}{\sum_{t=1}^{T} \hat{G}_t^{-1/2}}, \tag{10}$$

and the sample covariance matrix of $\hat{\epsilon}_t = \hat{G}_t^{-1}(\mathbf{x}_t - \hat{\mu}_X)$,

$$\hat{\Gamma} = \frac{\widehat{\mathbb{E}[G]}}{T} \sum_{t=1}^{T} \hat{\epsilon}_t \hat{\epsilon}_t^{\top}, \tag{11}$$

and take the eigenvalue decomposition, $\hat{\Gamma} = \hat{\mathbf{U}}\hat{\Lambda}\hat{\mathbf{U}}^{\top}$, which results in the updated estimator:

$$\tilde{\Gamma} = \hat{\mathbf{U}}\Lambda^{*}\hat{\mathbf{U}}^{\top}. \tag{12}$$

CM2-step: Given the CM1-step updates of μ_X, Γ, obtain a new update of θ_L by numerically maximizing the log-likelihood function $L_{\mathbf{X}}$.

Combining the ideas: ECME

Proposition

The above EM algorithm increases the likelihood $L_{\mathbf{X}}$ monotonically in each step.

Combining the ideas: Final Step

The last step of the algorithm is to obtain estimates of $\boldsymbol{\mu}$, **S** and **H** as,

$$\hat{\boldsymbol{\mu}} = \text{diag}(\hat{\Sigma})^{1/2}\hat{\boldsymbol{\mu}}_X + \bar{\mathbf{y}} \quad (13)$$

$$\hat{\mathbf{S}} = \text{diag}(\hat{\Sigma})/\widehat{\mathbb{E}[G]} \quad (14)$$

$$\tilde{\mathbf{H}} = \hat{\mathbf{S}}\tilde{\Gamma}\hat{\mathbf{S}}. \quad (15)$$

Combining the ideas: Final Step

- So far we developed an algorithm that is stable and fast, combining EM ideas with the method of moment principle.

Combining the ideas: Final Step

- So far we developed an algorithm that is stable and fast, combining EM ideas with the method of moment principle.
- However the monotonicity alone does not necessarily imply consistency of the parameter estimates.

Combining the ideas: Final Step

- So far we developed an algorithm that is stable and fast, combining EM ideas with the method of moment principle.
- However the monotonicity alone does not necessarily imply consistency of the parameter estimates.
- We now explore this in a simulation and empirical application.

Combining the ideas: Final Step

- So far we developed an algorithm that is stable and fast, combining EM ideas with the method of moment principle.
- However the monotonicity alone does not necessarily imply consistency of the parameter estimates.
- We now explore this in a simulation and empirical application.
- To do this we return to the MVG model with

$$G \sim \text{Gamma}(\lambda, 1).$$

1 Methodology

2 Simulation

3 Portfolio Application

4 References

Simulation

- There are two parameters of interest: $\mathbf{H}$ and λ.

Simulation

- There are two parameters of interest: $\mathbf{H}$ and λ.
- Settings: Number of Observations $T = 1'250$, $\lambda \in \{4, 8, 20\}$ and $N \in \{100, 500\}$.

Simulation

- There are two parameters of interest: $\mathbf{H}$ and λ.
- Settings: Number of Observations $T = 1'250$, $\lambda \in \{4, 8, 20\}$ and $N \in \{100, 500\}$.
- In all cases the true $\mathbf{H}$ is diagonal with
  1. 20% of eigenvalues are equal to 1,
  2. 40% of eigenvalues are equal to 3,
  3. 40% of eigenvalues are equal to 10,

  as in Ledoit and Wolf (2012, 2020a,b).

Simulation: Estimation of λ

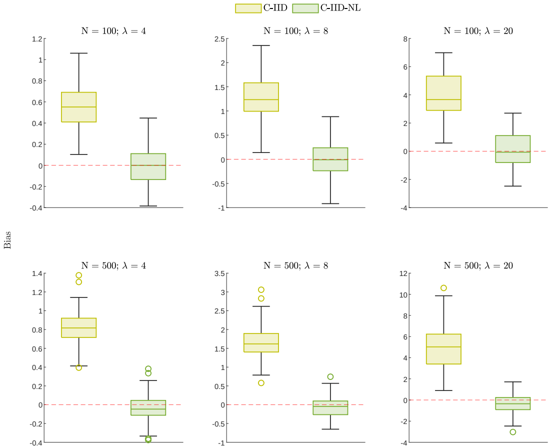

Figure 1: *Estimation accuracy of the two COMFORT (C) approaches for different values of* λ *and* N*. The number of observations is fixed to* $T = 1250$*. The simulation was carried out* $S = 100$ *times.*

Simulation: Estimation of H

To assess the estimation accuracy of $\mathbf{H}$, we study the so-called Percentage Relative Improvement (PRIAL), as in Engle et al. (2019),

$$100 \cdot \left(1 - \frac{\mathbb{E}[\hat{L}(\check{\mathbf{H}})]}{\mathbb{E}[\hat{L}(\hat{\mathbf{H}}_0)]} \right), \tag{16}$$

where

$$\hat{L}(\tilde{\mathbf{H}}) = \frac{\mathrm{Tr}(\tilde{\mathbf{H}}^{-1}\mathbf{H}\tilde{\mathbf{H}}^{-1}/N)}{\left[\mathrm{Tr}(\tilde{\mathbf{H}}^{-1})/N\right]^2} - \frac{1}{\mathrm{Tr}(\tilde{\mathbf{H}}^{-1})/N}, \tag{17}$$

with $\hat{\mathbf{H}}$ referring to the estimation of $\mathbf{H}$ obtained with the original algorithm, while $\tilde{\mathbf{H}}$ is obtained from the new Algorithm.

Simulation: Estimation of $\mathbf{H}$

Table 1: *PRIAL of COMFORT-IID-NL against the COMFORT-IID for different values of λ and N. The number of observations is fixed to $T = 1250$. The simulation was carried out $S = 100$ times.*

	$\lambda = 4$	$\lambda = 8$	$\lambda = 20$
$N = 100$	94.92	97.20	97.46
$N = 500$	96.20	98.64	100.69

1. Methodology
2. Simulation
3. Portfolio Application
4. References

Portfolio Application: Data and Portfolio Routine

- Based on the market capitalization, we consider daily observations of the $N \in \{100, 500, 1000\}$ largest stocks from CRSP (center for research in security prices).

Portfolio Application: Data and Portfolio Routine

- Based on the market capitalization, we consider daily observations of the $N \in \{100, 500, 1000\}$ largest stocks from CRSP (center for research in security prices).
- The out-of-sample period contains 480 months from 13.01.1981 until 31.12.2020.

Portfolio Application: Data and Portfolio Routine

- Based on the market capitalization, we consider daily observations of the $N \in \{100, 500, 1000\}$ largest stocks from CRSP (center for research in security prices).
- The out-of-sample period contains 480 months from 13.01.1981 until 31.12.2020.
- The estimation window is 1260 days and we rebalance every 21 days.

Portfolio Application: Data and Portfolio Routine

- Based on the market capitalization, we consider daily observations of the $N \in \{100, 500, 1000\}$ largest stocks from CRSP (center for research in security prices).
- The out-of-sample period contains 480 months from 13.01.1981 until 31.12.2020.
- The estimation window is 1260 days and we rebalance every 21 days.
- Global minimum variance portfolios (no short-sales constrains).

Portfolio Application: Data and Portfolio Routine

- Based on the market capitalization, we consider daily observations of the $N \in \{100, 500, 1000\}$ largest stocks from CRSP (center for research in security prices).
- The out-of-sample period contains 480 months from 13.01.1981 until 31.12.2020.
- The estimation window is 1260 days and we rebalance every 21 days.
- Global minimum variance portfolios (no short-sales constrains).
- The 9 out-of-sample portfolio statistics are:
  - average annualized return (*Average*)
  - annualized standard deviation (*Volatility*)
  - final cumulative return (*Total Return*)
  - maximum drawdown (*Max. Drawdown*)
  - annualized percentage turnover (*Turnover*)
  - annualized information ratio (*IR*)
  - annualized Sortino ratio (*Sortino*)
  - annualized percentage STARR-ratio at the 98.5% level ($STARR_{98.5\%}$)
  - annualized empirical expected shortfall at the 98.5% level ($ES_{98.5\%}$).

Portfolio Application: Summary Figure

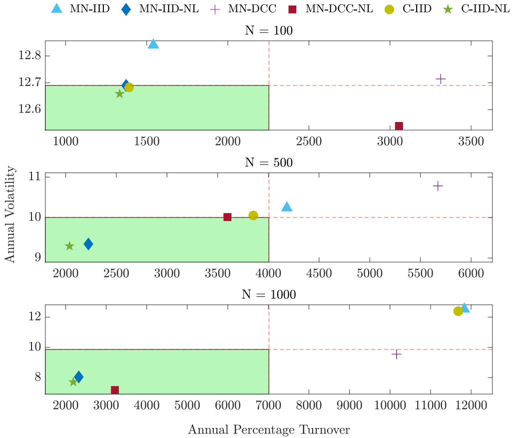

Figure 2: *For $N \in \{100, 500, 1000\}$ and the global minimum variance portfolio, the annual volatility is plotted against the annual percentage turnover. Each point in a subfigure represents one of the six models. The out-of-sample standard deviation is on the y-axis. The optimal region is highlighted in green (low turnover and low volatility).*

Portfolio Application: out-of-sample portfolio statistics

	Average	*Volatility*	*Total Return*	*Max. Drawdown*	*Turnover*	*IR*	*Sortino*	$STARR_{98.5\%}$	$ES_{98.5\%}$
				$N = 100$					
MN-IID	11.11	12.84	6003.55	−36.10	1540.84	0.87	1.22	3.88	286.19
MN-IID-NL	11.08	12.69	5980.10	−37.99	1373.08	0.87	1.23	3.91	283.19
MN-DCC	11.01	12.71	5800.18	−41.10	3312.61	0.87	1.21	3.87	284.77
MN-DCC-NL	11.19	12.54	6301.54	−41.40	3055.15	0.89	1.25	3.98	281.23
C-IID	10.90	12.68	5551.50	−35.03	1391.23	0.86	1.21	3.85	283.16
C-IID-NL	10.96	12.66	5710.96	−37.02	1333.05	0.87	1.22	3.87	282.99
				$N = 500$					
MN-IID	11.00	10.24	6483.87	−33.64	4182.63	1.07	1.52	4.74	231.97
MN-IID-NL	11.02	9.35	6769.77	−33.96	2224.79	1.18	1.64	5.23	210.47
MN-DCC	12.42	10.78	11320.10	−31.12	5672.42	1.15	1.72	5.72	217.19
MN-DCC-NL	12.77	10.01	13470.83	−25.58	3596.63	1.28	1.93	6.56	194.59
C-IID	10.90	10.05	6294.49	−32.30	3851.64	1.09	1.53	4.81	226.49
C-IID-NL	11.86	9.29	9558.15	−33.43	2038.75	1.28	1.80	5.79	204.84
				$N = 1000$					
MN-IID	12.19	12.53	9456.22	−29.60	11832.57	0.97	1.43	4.31	282.69
MN-IID-NL	12.02	8.04	10629.57	−27.57	2324.35	1.49	2.12	6.78	177.19
MN-DCC	12.65	9.55	13003.95	−32.98	10162.29	1.32	1.96	5.88	215.00
MN-DCC-NL	13.39	7.18	18958.32	−29.60	3215.62	1.86	2.74	8.99	148.89
C-IID	12.12	12.39	9261.59	−28.63	11682.46	0.98	1.44	4.35	278.60
C-IID-NL	12.89	7.71	15285.68	−26.33	2189.69	1.67	2.42	7.76	166.03

Table 2: *In the six model structures, MN stands for multivariate-Normal, C for COMFORT and NL for nonlinear shrinkage.*

1 Methodology

2 Simulation

3 Portfolio Application

4 References

Engle, R. F., Ledoit, O., and Wolf, M. (2019). Large Dynamic Covariance Matrices. *Journal of Business and Economic Statistics*, 37(2):363–375.

Ledoit, O. and Wolf, M. (2012). Nonlinear Shrinkage Estimation of Large-Dimensional Covariance Matrices. *The Annals of Statistics*, 40(2):1024 – 1060.

Ledoit, O. and Wolf, M. (2020a). Analytical Nonlinear Shrinkage of Large-Dimensional Covariance Matrices. *The Annals of Statistics*, 48(5):3043 – 3065.

Ledoit, O. and Wolf, M. (2020b). Quadratic Shrinkage for Large Covariance Matrices. *University of Zurich, Departmenf of Economics, Working Paper No. 335*.

Paolella, M. S. and Polak, P. (2015). COMFORT: A Common Market Factor Non-Gaussian Returns Model. *Journal of Econometrics*, 187(2):593–605.