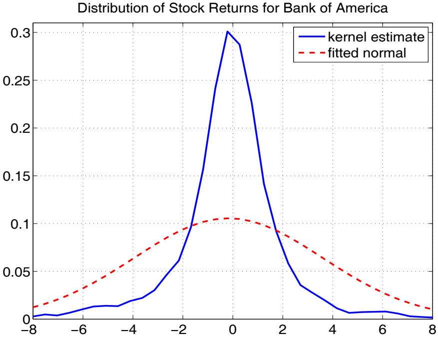
Multivariate Asset Return Prediction with Mixture Models

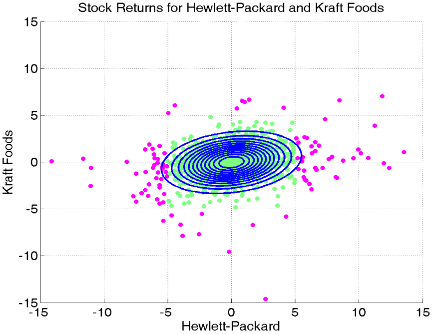
Marc S. Paolella

Swiss Banking Institute, University of Zürich

Introduction

The leptokurtic nature of asset returns has spawned an enormous amount of research into effective ways of modeling them, often involving the use of distributions which have power tails, and the ensuing nonexistence of higher moments.





Introduction

This issue has long since transcended ivory tower academics and quantitative groups in financial institutions, especially since the statement by Alan Greenspan (1997, p. 54),

> *... the biggest problems we now have with the whole evaluation of risk is the fat-tail problem, which is really creating very large conceptual difficulties. Because as we all know, the assumption of normality enables us to drop off a huge amount of complexity of our equations very much to the right of the equal sign. Because once you start putting in non-normality assumptions,* ***which unfortunately is what characterizes the real world****, then the issues become extremely difficult.*

Introduction

We propose a **multivariate** model which, besides exhibiting leptokurtic behavior and asymmetry,

- lends itself to tractable distributions of weighted sums of the marginal random variables (to enable portfolio construction),
- portfolios possess finite positive integer moments of all orders,
- is super-fast to estimate,
- delivers superior (compared to DCC and related models) multivariate density forecasts,
- has trivial stationarity conditions as a stochastic process. This is accomplished by **not using a GARCH structure**.

Need for Mixtures

A mixture can capture, besides excess kurtosis and asymmetry, the two further stylized facts of

- the *leverage* or *down market* effect—the negative correlation between volatility and asset returns, and
- *contagion effects*—the tendency of the correlation between asset returns to increase during pronounced market downturns, as well as during periods of higher volatility.

Other fat-tailed multivariate distributions, such as the multivariate Student's t or the multivariate generalized hyperbolic distribution, cannot achieve this.

Need for Mixtures

With a two-component multivariate normal mixture model with component parameters $\boldsymbol{\mu}_1$, $\boldsymbol{\Sigma}_1$, $\boldsymbol{\mu}_2$, $\boldsymbol{\Sigma}_2$, we would expect to have the primary component capturing the "business as usual" stock return behavior, with a near-zero mean vector $\boldsymbol{\mu}_1$, and the second component capturing the more volatile, "crisis" behavior, with

- (much) higher variances in $\boldsymbol{\Sigma}_2$ than in $\boldsymbol{\Sigma}_1$,
- significantly larger correlations, reflecting the contagion effect,
- and a predominantly negative $\boldsymbol{\mu}_2$, reflecting the down market effect.

A distribution with only a single mean vector and covariance (or, more generally, dispersion) matrix cannot capture this behavior, no matter how many additional shape parameters the distribution possesses.
This is true *even if each marginal were to be endowed with its own set of shape (tail and skew) parameters, as is possible when using a copula*.

Evidence for Mixtures

We use the 30 stocks comprising the DJIA-30, daily data from June 2001 to March 2009, for $T = 1945$ observations, on 30 variables.

Let $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,d})' \overset{\text{i.i.d.}}{\sim} \mathrm{Mix}_k\mathrm{N}_d(\mathbf{M}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, $t = 1, \ldots, T$, where $\mathrm{Mix}_k\mathrm{N}_d$ denotes the k-component, non-singular d-dimensional multivariate mixed normal distribution, with

$$\mathbf{M} = \left[\ \boldsymbol{\mu}_1 \mid \boldsymbol{\mu}_2 \mid \cdots \mid \boldsymbol{\mu}_k\ \right],\ \boldsymbol{\mu}_j = (\mu_{1j}, \mu_{2j}, \ldots, \mu_{dj})',\ \boldsymbol{\Psi} = \left[\ \boldsymbol{\Sigma}_1 \mid \boldsymbol{\Sigma}_2 \mid \cdots \mid \boldsymbol{\Sigma}_k\ \right],$$

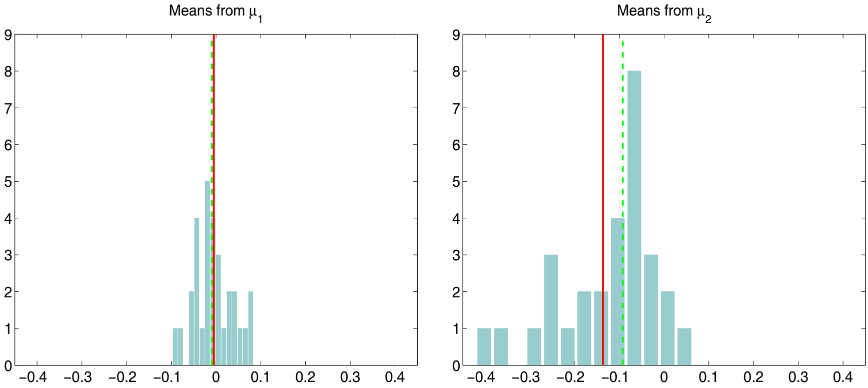
$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$, $\boldsymbol{\Sigma}_j > 0$ (i.e., positive definite), $j = 1, \ldots, k$, and

$$f_{\mathrm{Mix}_k\mathrm{N}_d}(\mathbf{y}; \mathbf{M}, \boldsymbol{\Psi}, \boldsymbol{\lambda}) = \sum_{j=1}^{k} \lambda_j f_{\mathrm{N}}\left(\mathbf{y}; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j\right), \quad \lambda_j \in (0,1), \quad \sum_{j=1}^{k} \lambda_j = 1,$$

with f_{N} denoting the d-variate multivariate normal distribution.

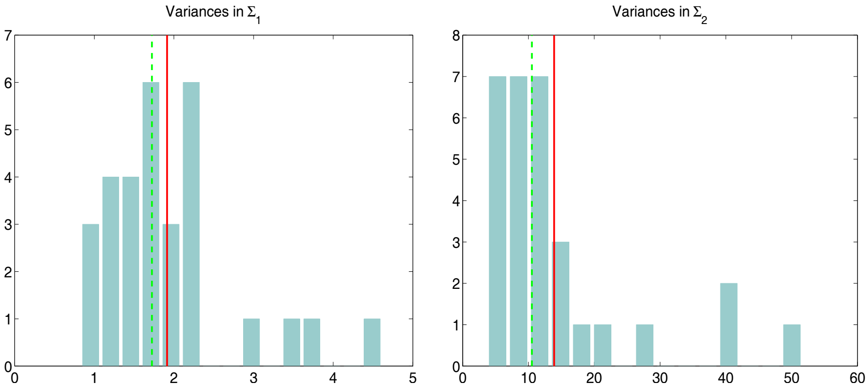
Evidence for Mixtures

For $k = 2$ components in the $\mathrm{Mix}_k\mathrm{N}_d$ distribution, show the $d = 30$ means for components 1 and 2 separately.



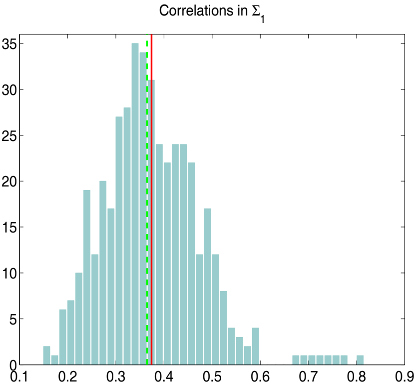
Evidence for Mixtures

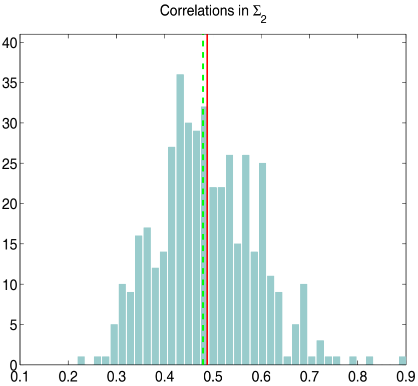
Same, but for the $d = 30$ variances for components 1 and 2 separately.



Evidence for Mixtures

Same, but for the $\binom{30}{2} = 425$ covariance for components 1 and 2 separately.





Shrinkage Estimation

Shrinkage estimation is crucial in this context for both numeric reasons, and (far) improved estimation.

The paper gives full details on how this is done.

Our shrinkage prior, as a function of the scalar hyper-parameter ω, is

$$a_1 = 2\omega, \quad a_2 = \omega/2, \quad c_1 = c_2 = 20\omega, \quad \mathbf{m}_1 = \mathbf{0}_p, \quad \mathbf{m}_2 = (-0.1)\mathbf{1}_p,$$
$$\mathbf{B}_1 = a_1\left[(1.5-0.6)\mathbf{I}_p + 0.6\mathbf{J}_p\right], \qquad \mathbf{B}_2 = a_2\left[(10-4.6)\mathbf{I}_p + 4.6\mathbf{J}_p\right]. \tag{1}$$

The only tuning parameter which remains to be chosen is ω.

The effect of different choices of ω is easily and informatively demonstrated with a **simulation study**, using the $\mathrm{Mix}_2\mathrm{N}_{30}$ model, with parameters given by the MLE of the 30 return series.

Shrinkage Estimation

For assessing the quality of the estimates, we use the log sum of squares as the summary measure, and noting that we have to convert the estimated parameter vector if the component labels are switched. That is,

$$M^*(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \min\left\{M(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}), M(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}^{=})\right\}, \tag{2}$$

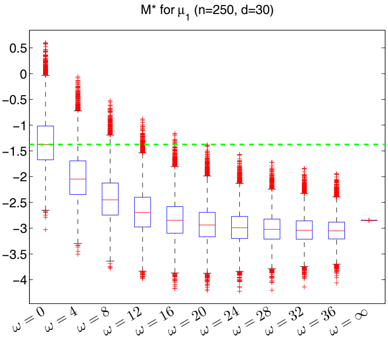
where $\boldsymbol{\theta} = \left(\boldsymbol{\mu}_1' \ \boldsymbol{\mu}_2' \ (\mathrm{vech}(\boldsymbol{\Sigma}_1))' \ (\mathrm{vech}(\boldsymbol{\Sigma}_2))' \ \lambda_1\right)'$, the vech operator of a matrix forms a column vector consisting of the elements on and below the main diagonal,

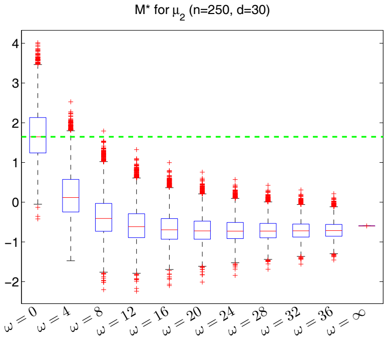
$$M(\widehat{\boldsymbol{\theta}}, \boldsymbol{\theta}) := \log(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})'(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \tag{3}$$

and $\boldsymbol{\theta}^{=}$ refers to the parameter vector obtained by switching the labels of the components, i.e., $\boldsymbol{\theta}^{=} = \left(\boldsymbol{\mu}_2' \ \boldsymbol{\mu}_1' \ (\mathrm{vech}(\boldsymbol{\Sigma}_2))' \ (\mathrm{vech}(\boldsymbol{\Sigma}_1))' \ (1-\lambda_1)\right)'$.

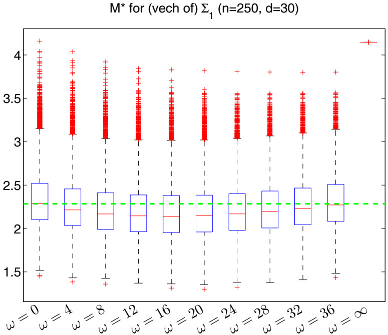
Shrinkage Estimation

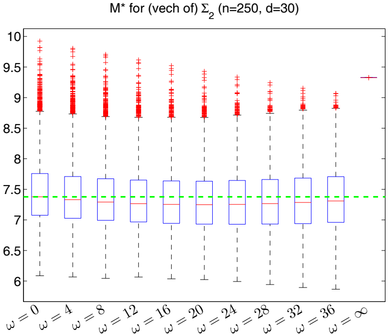
Estimation accuracy, measured as four divisions of M^* from (2) (λ_1 is ignored), based on simulation with 10,000 replications and $T = 250$, of the parameters of the $\mathrm{Mix}_2\mathrm{N}_{30}$ model, using as true parameters the MLE of the $d = 30$ return series, as a function of prior strength parameter ω.





Shrinkage Estimation

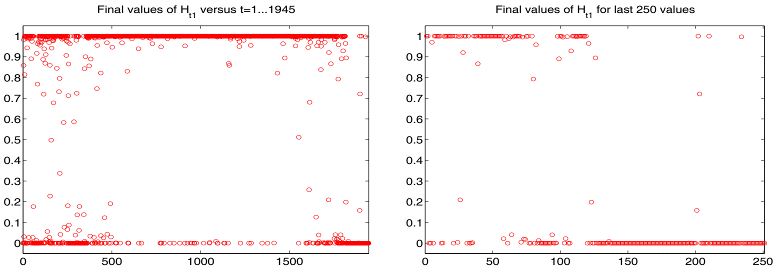




Component Separation

The $H_{t,j}$ are the posterior probabilities from the EM algorithm that observation $\mathbf{Y}_t$ came from component j, $t = 1, \ldots, T$, $j = 1, 2$, conditional on all the $\mathbf{Y}_t$ and the estimated components.

It is natural to plot these versus time.

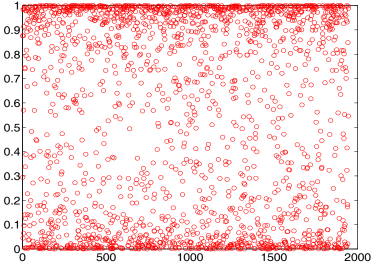


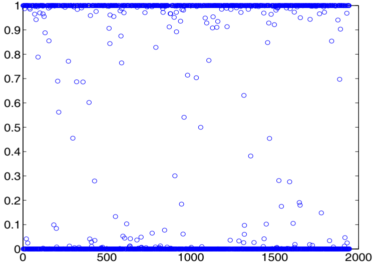
One might be tempted to use this finding as further evidence for the claim that there exist two, reasonably distinct, 'regimes' in financial markets, but this is not the case!

The same effect would occur if the data had arisen from a fat-tailed single-component multivariate distribution.

Component Separation

Now use a multivariate normal distribution (left) and a heavier-tailed multivariate Laplace, **both of which are single-component densities**! Then fit fit the 2-comp mixture of normals.





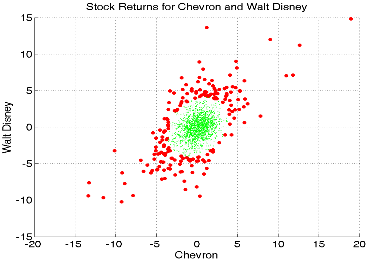
Thus, the separation apparent for the DJ-30 data is necessary, but clearly not sufficient to declare that the data were generated by a mixture distribution.

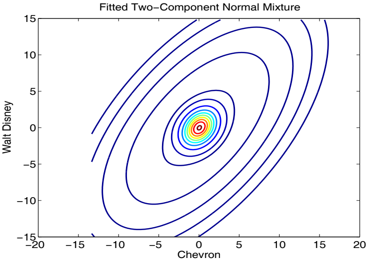
Stronger evidence for mixtures comes from the earlier graphic: the **means** of the first and second component differ markedly, with the latter being primarily negative, and the **correlations** in the second component are on average higher than those associated with component 1.

Further Evidence for Mixtures

Example: 2 stocks for which the correlations change the most between component 1 and 2.

For Chevron and Walt Disney, the component 1 correlation is 0.25; for component 2 it is 0.63.





Component Inspection: Is Normality Adequate?

The apparent separation is highly advantageous because it allows us to assign each $\mathbf{Y}_t$ to one of the two components with very high confidence.

Once done, we can assess how well each of the two estimated multivariate normal distributions fits the observations assigned to its component.

We use the criteria $H_{t,1} > 0.99$, choosing to place those $\mathbf{Y}_t$ whose corresponding values of $H_{t,1}$ suggest even a slight influence from component 2, into this more volatile component.

This results in 1,373 observations assigned to component 1, or 70.6% of the observations, and 572 to the second component.

Component Inspection: Is Normality Adequate?

Based on the split, for each of the $k \times d = 60$ series, we fit a flexible, asymmetric, fat-tailed distribution.
We use the generalized asymmetric t, or GAt, distribution, with location-zero, scale-one pdf given by $(p, \nu, \theta \in \mathbb{R}_{>0})$

$$f_{\mathrm{GAt}}(z; p, \nu, \theta) = C_{p,\nu,\theta} \times \begin{cases} \left(1 + \dfrac{(-z \cdot \theta)^p}{\nu}\right)^{-\left(\nu + \frac{1}{p}\right)}, & \text{if } z < 0, \\ \left(1 + \dfrac{(z/\theta)^p}{\nu}\right)^{-\left(\nu + \frac{1}{p}\right)}, & \text{if } z \geq 0, \end{cases}$$

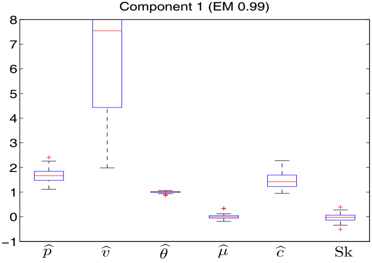
Parameter p measures the peakedness of the density, with values near one indicative of Laplace-type behavior, while values near two indicate a peak similar to that of the Gaussian and Student's t distributions.

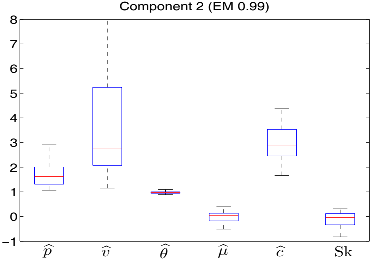
Parameter v indicates the tail thickness, and is analogous to the degrees of freedom parameter.

Parameter θ, controls the asymmetry, with values less than 1 indicating negative skewness.

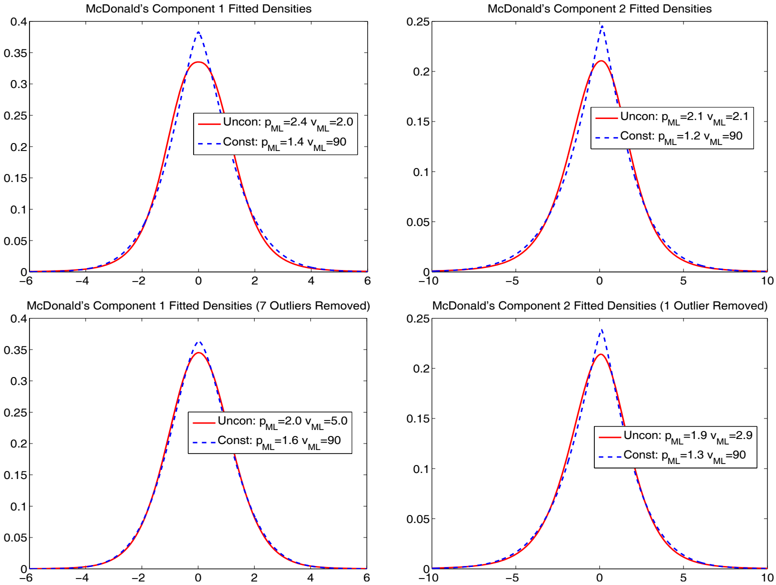
Component Inspection: Is Normality Adequate?

Remember: Parameter v indicates the tail thickness analogous to the df in the Student't t, except that moments of order $v \cdot p$ and higher do not exist. So that, if $p = 2$, then we would double the value of v to make it comparable to the df in the usual Student's t.





Worst Case: McDonalds



Component Inspection: Is Normality Adequate?

For each of the 30 series, but *not* separating them into the two components, we fit the GAt, with no parameter restrictions, and with the restriction that $90 < \hat{v} < 100$ (forces normality if $p = 2$ and $\theta = 1$, or Laplace if $p = 1$ and $\theta = 1$)

Compute the asymptotically valid p-value of the likelihood ratio test. If that value is less than 0.05, then we remove the largest value (in absolute terms) from the series, and re-compute the estimates and the p-value. This is repeated until the p-value exceeds 0.05.

We report the smallest number of observations required to be removed in order to achieve this.

Do this for the actual 30 series, and also the two separated components.

Component Inspection: Is Normality Adequate?

Stock number 5 is Bank of America, and shows that 65 most extreme values had to be removed from the series to get the *p*-value above 0.05, but **no observations** from component 1 needed to be removed, and only 3 from component 2.

Stock #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
All	19	19	5	2	65	21	11	27	60	10	30	19	28	31	8
Comp1	0	0	0	0	0	3	0	0	1	1	1	0	0	0	0
Comp2	0	1	0	0	3	1	1	6	8	2	0	8	0	1	1

Stock #	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
All	7	12	8	5	19	7	15	28	3	14	10	11	6	8	10
Comp1	0	0	2	2	0	0	7	0	0	0	0	0	0	0	0
Comp2	0	0	0	1	0	0	1	2	0	2	1	3	0	2	1

Density Forecasting

- We forecast the entire multivariate density.
- The measure of interest is what we will call the **(realized) predictive log-likelihood**, given by

$$\pi_t(\mathcal{M}, v) = \log f^{\mathcal{M}}_{t|I_{t-1}}(\mathbf{y}_t; \widehat{\psi}), \tag{4}$$

  where v denotes the size of the rolling window used to determine I_{t-1} (and the set of observations used for estimation of ψ) for each time point t.

Density Forecasting

- We suggest to use what we refer to as the **normalized sum of the realized predictive log-likelihood**, given by

$$S_{\tau_0,T}(\mathcal{M}, v) = \frac{1}{(T - \tau_0)\, d} \sum_{t=\tau_0+1}^{T} \pi_t(\mathcal{M}, v), \tag{5}$$

  where d is the dimension of the data.
- It is thus the average realized predictive log-likelihood, averaged over the number of time points used and the dimension of the random variable under study. This facilitates comparison over different d, τ_0 and T.
- In our setting, we use the $d = 30$ daily return series of the DJ-30, with $v = \tau_0 = 500$, which corresponds to two years of data, and $T = 1,945$.

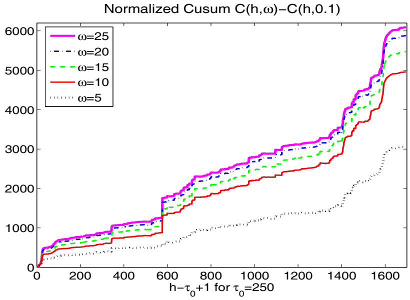
Density Forecasting

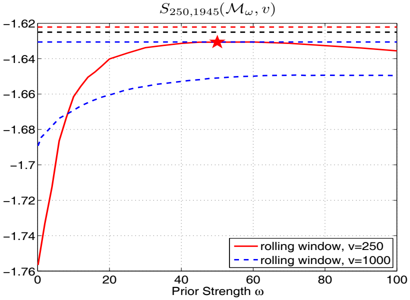
We compute for set of models indexed by ω, $\pi_t(\mathcal{M}_\omega, v)$ from (4).

We take $\{\mathcal{M}_\omega\}$ to be the $\mathrm{Mix}_2\mathsf{N}_{30}$ model estimated with shrinkage prior (1) for a given value of hyper-parameter ω.

We do this using a moving window of size $v = 250$, starting at observation $\tau_0 = v = 250$, and updating parameter vector $\widehat{\boldsymbol{\theta}}$ at each time increment.

Density Forecasting: Finding the optimal ω

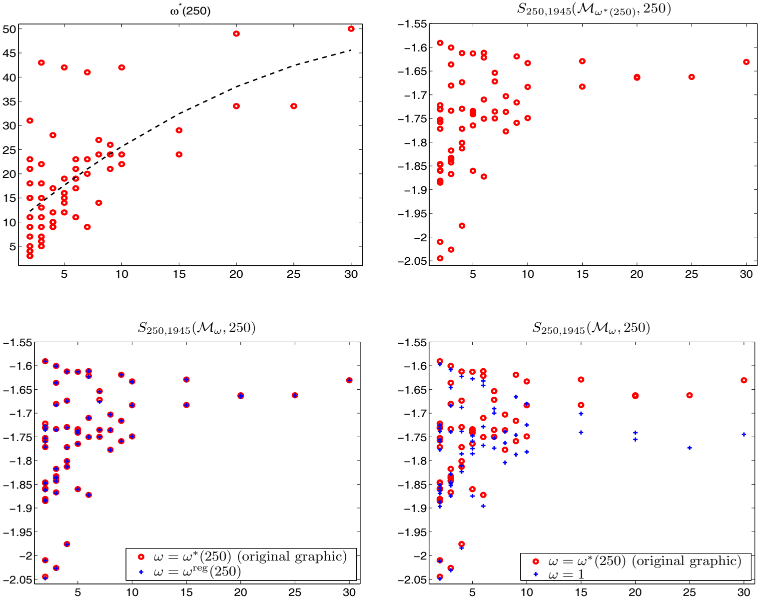




Here,

$$C(h, \omega) = \sum_{t=\tau_0+1}^{h} \pi_t(\mathcal{M}_\omega, 250), \quad h = \tau_0 + 1, \ldots, T,$$

Optimal Shrinkage as a Function of d

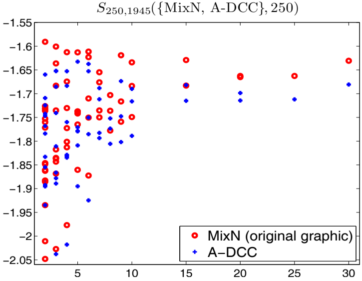


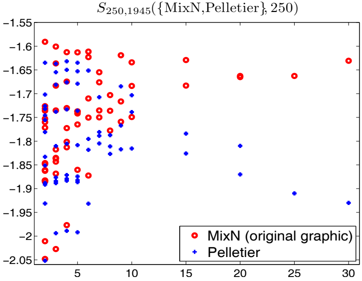
The Competition: CCC, DCC, A-DCC, RSDC-GARCH

It turns out that the forecasted values yield only a barely visible preference of DCC over CCC, and A-DCC over DCC.

In light of what appears to be a general consensus in the literature that DCC is significantly superior to CCC, and the encouraging in-sample results favoring asymmetry, these results were somewhat surprising.

The Regime Switching Dynamic Correlation (RSDC-) GARCH model of Pelletier (2006) also recognizes different "regimes" in the data.

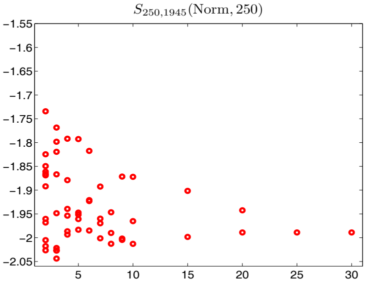


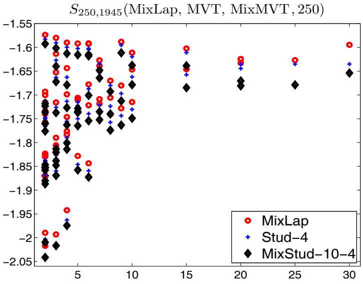


The Competition: MVN and MVT

Use of the MVN in the left panel. Just as a benchmark...

Right panel shows: (i) the multivariate Student's t distribution (MVT) with fixed degrees of freedom 4 (ii), the two-component multivariate Student t mixture (using 10 and 4 degrees of freedom, for the two components), and (iii) the $\mathrm{Mix}_2\mathrm{Lap}_d$ distribution introduced below.





Why does $\mathrm{Mix}_2\mathrm{Lap}_d$ beat $\mathrm{Mix}_2\mathrm{Stud}_d$

From a forecasting point of view, the Laplace mixture resulted in superior performance. We conjecture the reason for this to be that the tails of the t, *when used in a mixture context*, are too fat.

While higher kurtosis is indeed still required for the second component, the Laplace offers this without power tails and the ensuing upper bound on finite absolute positive moments.

Moreover, as demonstrated above via use of the GAt distribution and the fitted values of peakedness parameter p, there is evidence that the two components, particularly the second, have a more peaked distribution than the Student's t.

Portfolio Construction

Let $\mathbf{Y} \sim \mathrm{Mix}_k\mathrm{N}_d(\mathbf{M}, \boldsymbol{\Psi}, \boldsymbol{\lambda})$, with, as before,
$\mathbf{M} = \left[\ \boldsymbol{\mu}_1 \mid \boldsymbol{\mu}_2 \mid \cdots \mid \boldsymbol{\mu}_k\ \right]$, $\boldsymbol{\Psi} = \left[\ \boldsymbol{\Sigma}_1 \mid \boldsymbol{\Sigma}_2 \mid \cdots \mid \boldsymbol{\Sigma}_k\ \right]$, and
$\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_k)$.

Interest centers on the distribution of the portfolio $P = \mathbf{a}'\mathbf{Y}$, $\mathbf{a} \in \mathbb{R}^d$.

We show in the Appendix that

$$f_P(x) = \sum_{c=1}^{k} \lambda_c \phi\left(x; \mathbf{a}'\boldsymbol{\mu}_c, \mathbf{a}'\boldsymbol{\Sigma}_c\mathbf{a}\right), \tag{6}$$

where $\phi\left(x; \mu, \sigma^2\right)$ denotes the normal distribution with mean μ and variance σ^2 evaluated at x.

With such a simple analytic result, portfolio optimization is straightforward, and simulation (as is required when copula-based methods are used) is not necessary.

Portfolio Construction

With $\mu_c = \mathbf{a}'\boldsymbol{\mu}_c$ and $\sigma_c^2 = \mathbf{a}'\boldsymbol{\Sigma}_c\mathbf{a}$, $c = 1, \ldots, k$,

$$\mu_P = \mathbb{E}[P] = \sum_{c=1}^{k} \lambda_c \mu_c, \qquad \sigma_P^2 = \mathbb{V}(P) = \sum_{c=1}^{k} \lambda_c \left(\sigma_c^2 + \mu_c^2\right) - \mu_P^2.$$

It is common when working with non-Gaussian portfolio distributions to use a measure of downside risk, such as the value-at-risk (VaR), or the expected shortfall (ES).

Portfolio Construction

The VaR involves the γ-quantile of P, denoted $q_{P,\gamma}$, for $0 < \gamma < 1$, with γ typically 0.01 or 0.05, and which can be found numerically by using the cdf of P, easily seen to be $F_P(x) = \sum_{c=1}^{k} \lambda_c \Phi\left((x - \mu_c)/\sigma_c\right)$, with Φ the standard normal cdf.

The γ-level ES of P is given by

$$\mathrm{ES}_\gamma(P) = \frac{1}{\gamma} \int_0^\gamma Q_P(p)\,\mathrm{d}p,$$

where Q_P is the quantile function of P. We prove in the Appendix that this integral can be represented analytically as

$$\mathrm{ES}_\gamma(P) = \sum_{j=1}^{k} \frac{\lambda_j \Phi(c_j)}{\gamma} \left\{ \mu_j - \sigma_j \frac{\phi(c_j)}{\Phi(c_j)} \right\}, \quad c_j = \frac{q_{P,\gamma} - \mu_j}{\sigma_j}, \quad q_{P,\gamma} = Q_P(\gamma).$$

Multivariate Laplace

The d-variate multivariate Laplace distribution is given by

$$f_{\mathbf{Y}}(\mathbf{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, b) = \frac{1}{|\boldsymbol{\Sigma}|^{1/2} (2\pi)^{d/2}} \frac{2}{\Gamma(b)} \left(\frac{m}{2}\right)^{b/2-d/4} K_{b-d/2}\left(\sqrt{2m}\right), \quad (7)$$

where $m = (\mathbf{y} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \boldsymbol{\mu})$.

It generalizes several constructs in the literature, but itself is a special case of the multivariate generalized hyperbolic distribution.

We write $\mathbf{Y} \sim \mathrm{Lap}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, b)$.

Mixtures of Multivariate Laplace

We say the d-dimensional random variable $\mathbf{Y}_i$ follows a k-component mixture of multivariate Laplace distributions, or $\mathrm{Mix}_k\mathrm{Lap}_d$, if its distribution is given by

$$f_{\mathrm{Mix}_k\mathrm{Lap}_d}\left(\mathbf{y};\mathbf{M},\mathbf{\Psi},\boldsymbol{\lambda},\mathbf{b}\right)=\sum_{j=1}^{k}\lambda_j f_{\mathrm{Lap}}\left(\mathbf{y};\boldsymbol{\mu}_j,\mathbf{\Sigma}_j,b_j\right),$$

$\lambda_j\in(0,1)$, $\sum_{j=1}^{k}\lambda_j=1$, with f_{Lap} denoting the d-variate multivariate Laplace distribution given above, $\mathbf{M}$, $\mathbf{\Psi}$ and $\boldsymbol{\lambda}$ are as with the $\mathrm{Mix}_k\mathrm{Norm}_d$, and $\mathbf{b}=(b_1,\ldots,b_k)'$.

Mixtures of Multivariate Laplace

We observe d-variate random variables $\mathbf{Y}_t = (Y_{t,1}, Y_{t,2}, \ldots, Y_{t,d})'$, $t = 1, \ldots, T$, with $\mathbf{Y}_t \overset{\text{i.i.d.}}{\sim} \mathrm{Mix}_k\mathrm{Lap}_d(\mathbf{M}, \boldsymbol{\Psi}, \boldsymbol{\lambda}, \mathbf{b})$, and take the values of $\mathbf{b} = (b_1, \ldots, b_k)$ to be known constants.

Interest centers on estimation of the remaining parameters.

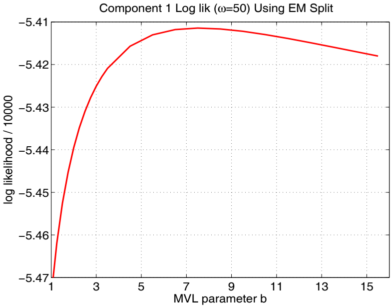
This is conducted via an EM algorithm, the derivation of which is given in the Appendix.

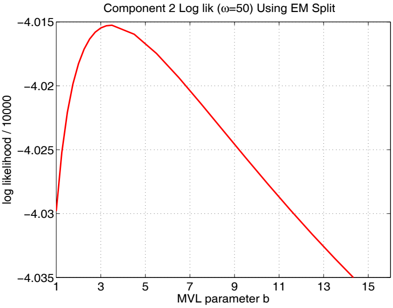
In addition, the EM recursions are extended there to support use of a quasi-Bayesian paradigm analogous to that used for the $\mathrm{Mix}_k\mathrm{N}_d$ distribution.

Estimation of the b_i in the Mixture of Multivariate Laplace

We use the profile likelihood to each component, applied to the split data via the EM algorithm.

The details are in the paper. This is not optimal, but adequate.





Portfolio Construction for $\mathrm{Mix}_k\mathrm{Lap}_d$

Let $\mathbf{L} \sim \mathrm{Lap}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}, b\right)$ with density (7).

Then, for $\mathbf{a} \in \mathbb{R}^d$, $P = \mathbf{a}'\mathbf{L} \sim \mathrm{Lap}\left(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}', b\right)$, which is a special case of the general result for normal mixture distributions, as shown in, e.g., McNeil, Embrechts and Frey (2005, p. 76).

Now let $\mathbf{Y} \sim \mathrm{Mix}_k\mathrm{Lap}_d\left(\mathbf{M}, \boldsymbol{\Psi}, \boldsymbol{\lambda}, \mathbf{b}\right)$ and $P = \mathbf{a}'\mathbf{Y}$.

Then, analogous to the Mix MVN case, and using the same format of proof, we find that

$$f_P\left(x\right) = \sum_{c=1}^{k} \lambda_c \,\mathrm{Lap}\left(x; \mathbf{a}'\boldsymbol{\mu}_c, \mathbf{a}'\boldsymbol{\Sigma}_c\mathbf{a}, b_c\right).$$

Portfolio Construction for $\mathrm{Mix}_k\mathrm{Lap}_d$

Similar to (33), we have, with $\mu_c = \mathbf{a}'\boldsymbol{\mu}_c$ and $\sigma_c^2 = \mathbf{a}'\boldsymbol{\Sigma}_c\mathbf{a}$, $c = 1, \ldots, k$,

$$\mu_P = \mathbb{E}[P] = \sum_{c=1}^{k} \lambda_c \mu_c, \qquad \sigma_P^2 = \mathbb{V}(P) = \sum_{c=1}^{k} \lambda_c \left(b_c \sigma_c^2 + \mu_c^2 \right) - \mu_P^2.$$

A closed-form expression for the expected shortfall is currently not available.

Though given the tractable density function of P, and the exponential (not power) tails of the Laplace distribution, numeric integration to compute the relevant quantile, and the integral associated with the expected shortfall, will be fast and reliable.